Henckens *et al. Art Int Surg.* 2025;5:36-45
DOI: 10.20517/ais.2024.47

Artificial Intelligence Surgery

Guideline

Assessment of the extent of lymphadenectomy in esophageal cancer surgery in the observational TIGER study: [TIGER-SQA] study protocol

Sofie P.G. Henckens[1,2], Nannet Schuring[1,2], Dillen C. van der Aa[1,2], Alexander Harris[3,4], Stella Mavroveli[3], Alexander A.J. Grüter[2,5], Martin Wagner[6,7,8,9], Marie Daum[6], Ivana Išgum[2,10,11], Sebastian Bodenstedt[9,12], Stefanie Speidel[9,12], Beat P. MüllerStich[13], Sheraz R. Markar[14], Mark I. van Berge Henegouwen[1,2], George B. Hanna[3], Suzanne S. Gisbertz[1,2]

[1]Department of Surgery, Amsterdam UMC, location University of Amsterdam, Amsterdam 1105 AZ, the Netherlands.
[2]Cancer Center Amsterdam, Cancer Treatment and Quality of Life, Amsterdam, the Netherlands.
[3]Department of Surgery and Cancer, Imperial College London, London SW7 2AZ, UK.
[4]Department of Upper Gastrointestinal Surgery, Barking Havering and Redbridge University Hospitals NHS Trust, Queen's Hospital, Romford RM7 0AG, UK.
[5]Department of Surgery, Amsterdam UMC, location Vrije Universiteit Amsterdam, Amsterdam 1081 HV, the Netherlands.
[6]Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden 01307, Germany.
[7]Department of Visceral-, Thoracic and Vascular Surgery, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden 01307, Germany.
[8]National Center for Tumor Diseases (NCT), Heidelberg 69120, Germany.
[9]Center for Tactile Internet with Human-in-the-loop (CeTI), TU Dresden, Dresden 01069, Germany.
[10]Department of Biomedical Engineering and Physics, Amsterdam UMC, location University of Amsterdam, Amsterdam 1105 AZ, the Netherlands.
[11]Department of Radiology and Nuclear Medicine, Amsterdam UMC, location University of Amsterdam, Amsterdam 1105 AZ, the Netherlands.
[12]Department of Translational Surgical Oncology, National Center for Tumor Diseases Partner Site Dresden, Dresden 01307, Germany.
[13]Department for Abdominal Surgery, Clarunis-University Digestive Health Care Center, St. Clara Hospital and University Hospital Basel, Basel 4031, Switzerland.
[14]Department of Surgery, Churchill Hospital, Oxford University Hospitals NHS Trust, Oxford OX3 7LE, UK.

Correspondence to: Sofie P.G. Henckens, Department of Surgery, Amsterdam UMC, location University of Amsterdam, Amsterdam 1105 AZ, the Netherlands. E-mail: s.p.henckens@amsterdamumc.nl

How to cite this article: Henckens SPG, Schuring N, van der Aa DC, Harris A, Mavroveli S, Grüter AAJ, Wagner M, Daum M, Išgum I, Bodenstedt S, Speidel S, MüllerStich BP, Markar SR, van Berge Henegouwen MI, Hanna GB, Gisbertz SS. Assessment of the extent of lymphadenectomy in esophageal cancer surgery in the observational TIGER study: [TIGER-SQA] study protocol. *Art Int Surg.* 2025;5:36-45. https://dx.doi.org/10.20517/ais.2024.47

Received: 21 Jun 2024 **First Decision:** 27 Aug 2024 **Revised:** 28 Nov 2024 **Accepted:** 10 Dec 2024 **Published:** 8 Jan 2025

Academic Editor: Andrew Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

The distribution pattern of esophageal cancer lymph node metastases and consequently the lymph node stations that should be resected are currently being investigated in the international distribution of lymph node metastases in esophageal carcinoma (TIGER) study. This observational study has no specific entry criteria, resulting in significant variation among participating centers regarding the extent of lymphatic tissue clearance during surgery. To ensure reliable interpretation of the TIGER study results, this study aims to develop an intraoperative metric (surgical quality assessment tool) to examine the extent of lymphadenectomy. This multicenter prospective study assesses the extent of lymphadenectomy during esophagectomies performed within the TIGER study. After consensus on the applicability of surgical quality assessment in observational studies (phase 1), a short photo/video of the thoracic, abdominal, and [if applicable] cervical area will be captured after lymphadenectomy (phase 2). Those images will be rated by expert surgeons using a structured assessment tool (phase 3), and an automatic artificial intelligence (AI)-based quality assessment will be developed (phase 4). The variability in lymphadenectomy will be adjusted when analyzing TIGER study outcomes. Technique standardization by surgical quality assurance (SQA) reduces variation in trial outcomes. This cannot be applied to the TIGER study because of its observational nature, although surgical heterogeneity might influence TIGER study outcomes. To monitor surgical performance within observational studies, surgical quality assessment can be applied, although this has not been done yet. This study will ensure a reliable interpretation of TIGER study results and could be used as a benchmark for quality assessment in future surgical studies.

Keywords: Esophagectomy, lymph node metastases, lymphadenectomy, surgical variability, surgical quality assessment, artificial intelligence

INTRODUCTION

Esophageal carcinoma affects half a million new patients annually and is the seventh most common cancer in the world[1]. After multimodal treatment involving neoadjuvant chemo(radio)therapy and surgical resection of the esophagus, 5-year survival is approximately 50%[2,3]. The lymphatic system surrounding the esophagus is complexly organized, contributing to the multidirectional spread of metastatic cells to lymph nodes[4]. Therefore, one important prognostic parameter for esophageal cancer patients is lymph node status[5,6].

The exact distribution pattern of esophageal lymph node metastases is, however, unpredictable and depends on multiple factors such as tumor location, histology, tumor biology, invasion depth, and neoadjuvant treatment[7-12]. Because of the remaining uncertainty on the specific pattern of lymphatic spread, this is being investigated in the prospective observational multicenter distribution of lymph node metastases in esophageal carcinoma (TIGER) study[13]. The TIGER study will eventually identify the lymph node stations that should be resected in relation to patient, tumor and treatment characteristics, and will establish patient-tailored treatment in which the extent of lymphadenectomy during surgery is adapted according to the pattern of lymph node metastases of the individual patient.

However, while the TIGER study investigates the distribution pattern of lymph node metastases, centers participating in the study show great disparity in the extent to which structures are being cleared of lymphatic tissue[14]. This results from the above-described lack of current knowledge, the absence of consensus on the optimal lymphadenectomy extent, and the observational nature of the TIGER study without a prescribed extent of lymphadenectomy. This variability in the extent of lymphadenectomy might compromise the generalizability and external validity of TIGER study results. Studies examining the prognostic value of lymphadenectomy often use traditional proxies such as lymph node yield to quantify the extent of lymphadenectomy, as this is what is collected during surgery. However, providing an actual intraoperative metric, by quantifying what happens during surgery rather than objectifying the quantity of the lymphadenectomy based on pathological proxies, could be more reflective. Essentially, the value of solid organ cancer surgery is what is left behind in the patient, and thus, we must find a new metric to intraoperatively assess surgical performance directly. Furthermore, pathology findings are limited by the surgical approach, as it is impossible for the pathologist to identify any metastatic involvement in lymph node stations that are not surgically dissected. As a result, this leads to an inaccurately represented distribution of lymph node metastases, potentially underestimating the extent of disease involvement.

Standardization of surgery by surgical quality assurance (SQA) has been shown to reduce variation in outcomes in randomized controlled trials (RCTs) and it is, therefore, increasingly being implemented in trials[15-17]. An alternative, more suitable SQA variant for observational studies (surgical quality assessment), is presumably also needed to control the heterogeneity in surgical performances, although this has never been investigated before. To this end, the current study, TIGER-SQA, has been designed to ensure a reliable interpretation of TIGER results according to the actual lymphadenectomy performed during each operation. The SQA approach in the observational TIGER study is different from SQA methods in interventional RCTs, whereby SQA is used to standardize the intervention and to act as a gatekeeper for entry into the trial. The aim of this TIGER-SQA study is to develop and utilize a surgical quality assessment tool to examine the extent of lymphadenectomy during esophagectomy for esophageal carcinoma, within the prospective, observational, multicenter TIGER study.

METHODS

Study design and setting

This quality assessment study (TIGER-SQA) is a prospective, observational, multicenter study to assess the extent of lymphatic tissue removal during esophagectomy for esophageal cancer.

Study objectives

The objectives of this TIGER-SQA study are:

1. To reach a consensus on how to best apply SQA in observational studies, and to develop a detailed photo/video SQA tool and assessment structure.
2. To capture the different operative fields of patients included in the TIGER study after completion of lymphadenectomy.
3. To assess the extent of the removal of lymph node stations by expert surgeons using the developed photo/video SQA tool.
4. To assess the extent of the removal of lymph node stations also by an artificial intelligence (AI)-based automatic quality assessment system.

Patient population

All patients included in the TIGER study from September 2022 are eligible for this TIGER-SQA study. The TIGER study population consists of patients with a resectable (cT1-4a, N0-3, M0) esophageal or gastro-esophageal junction carcinoma, in whom a transthoracic esophagectomy with 2- or 3-field lymphadenectomy is performed. A detailed description of the in- and exclusion criteria can be found in the previously published TIGER study protocol[13]. Preoperative staging, treatment, and pathology modalities will generally be performed according to relevant national guidelines and may differ per country.

Study phases

Phase 1

First, experts will reach a consensus on the proposed implementation of SQA within the scope of the TIGER study through four virtual focus group discussions. The participants of the focus groups will be purposefully selected expert surgeons and pathologists on esophageal cancer, with representatives from all nineteen countries participating in the TIGER study. During the focus groups, we will gain expert opinion on the applicability of SQA in the TIGER study, and discuss the differences between SQA within RCTs *vs.* SQA within observational studies. Furthermore, we will discuss the preferences of using either photographs and/or video snapshots to adequately evaluate the extent of lymphadenectomy. During these meetings, we will also seek expert opinions on the optimal approach for scoring lymphadenectomy to guide the development of the SQA tool.

Subsequently, an existing photo/video SQA tool, designed by members of this research group, will be further refined for assessment of 2- and 3-field lymphadenectomy during esophagectomy[17]. This tool has been used previously in the ROMIO trial, in which it has shown to be reliable and acceptable for SQA of 2-stage esophagectomy for surgical oncology RCTs by several expert surgeons. Adapting the existing tool is necessary because the current tool focuses on esophagectomy in general, not specifically on lymphadenectomy. In addition, the existing tool does not include lymph node stations in the upper mediastinum (for complete mediastinal dissection) and the neck (3-field dissection). In terms of content validity, the developed tool should comprehensively cover all aspects of lymphadenectomy that are critical for assessing its adequacy. Specifically, it should include 19 items to rate, representing all stations of the TIGER lymph node classification. Additionally, a detailed description of anatomical boundaries for each station is essential to make sure individual assessors use the same criteria when deciding whether a particular station has been cleared of lymph nodes, thus minimizing inter-rater variability. Anatomical boundaries per lymph node station must be an evidence-based framework based on expert surgeons' opinions and existing esophageal lymph node classifications. Inter-rater variability in assessments will further be minimized by a tool with clear, not too many, answer options, and clear instructions on how a complete lymphadenectomy should be performed per station to be assessed as complete lymphadenectomy. The tool needs to be straightforward and easy to use, requiring minimal time and effort from the surgeon to provide assessments.

Next, a framework for the photo/video assessment methodology will be developed. Each lymph node station of the TIGER study lymph node classification, will be marked as (a) complete; (b) incomplete; (c) not performed; (d) insufficient evidence to provide a rating (blurred photographs or obstructed field of view); or (e) absent data (no video submitted)[18]. A "complete" lymphadenectomy is defined as: all lymphatic tissue in the designated station has been fully removed and anatomical landmarks are clearly visible. When severe lymph nodes are still visible and anatomical landmarks are not fully exposed, the procedure should be rated as "incomplete". If the lymphatic tissue appears unchanged, indicating that no lymphadenectomy had been performed, the station should be rated as "not performed". The final two categories address situations where either the submitted photo or video is blurred or the field of view is obstructed, making a conclusive assessment impossible, or where no data was submitted at all.

After completion of the above, a PILOT study will be conducted to test the developed tool and assessment framework. This study will use a subset of video images and involve a small group of raters (selected from the larger pool of designated esophagogastric surgeons who will also rate later in the study). The final exact framework for the photo/video assessment is an end product of phase 1 of this study and will be discussed in a future manuscript. The PILOT study will give final answers on whether the tool is clear and complete, on a clear definition of complete lymphadenectomy, and on how the rating will be performed exactly. The planned PILOT will also tell whether an assessor training workshop is necessary to align the assessors in using the tool, which is the case when the PILOT study delivers a lot of heterogeneity in ratings.

Phase 2

The second phase of the study includes the utilization of the developed photo/video SQA tool within the TIGER study. As the primary endpoint of the TIGER study is the distribution of lymph node metastases, a tool that reflects the quality of the end product (lymphadenectomy) is essential. To achieve this, photos/videos of each anatomical region [thoracic, abdominal and (if applicable) cervical] will be taken at the end of the lymphadenectomy. These images must clearly capture the operative fields after completion of the lymphadenectomy, showing the skeletonization of all the designated landmarks. To ensure short, equal videos of high quality, we will develop a video manual for the participating surgeons, outlining the exact anatomical landmarks that should be exposed. For minimally invasive esophagectomy, laparoscopic (robotic or manual) images will be captured, while for open esophagectomy, the images will be captured by a video camera. The participating center will upload the images to a secured digital platform (FormsVision, Alea Clinical), and a notification will be sent to one of the raters once the case is ready for assessment.

Phase 3

The rater will assess the extent of the actual lymphadenectomy performed as shown in the photo/video, using the SQA tool incorporated into the digital platform. The raters will not know the surgeon or the center from which the photos/videos were obtained. The rating of images of each case will be randomly undertaken by one of the designated esophagogastric surgeon assessors, all of whom are part of the TIGER steering committee and experienced in performing > 100 esophagectomies. A senior experienced assessor will additionally assess a randomly selected 5% of all cases. If the senior assessor disagrees with the rating of the primary assessor, they will discuss the case and come to an agreement. The planned PILOT study will determine whether the preselected structure and the choice for only one assessor is appropriate or whether it results in excessive inter-rater reliability, necessitating the assessment of each video by multiple raters and/or an increase in the percentage of double-rated assessments.

Phase 4

Alongside this human assessment by expert surgeon raters, we propose an assessment of the photos/videos for the extent of lymphadenectomy by automatic AI-based surgical assessment by computers. Since AI-guided video assessment on the completeness of lymphadenectomy in esophagectomy is a novel approach, this project will be established from the ground up for this specific purpose of the TIGER-SQA study. Consequently, an annotation protocol must first be developed, and part of the photos/videos captured and collected in phase 2 will be annotated accordingly to train the deep learning algorithm. The algorithm will be trained primarily to recognize the anatomy of the abdomen and thorax, including anatomical structures, their boundaries, and lymphatic tissue, and it will learn to classify tissues, such as aortic, lung, or lymphatic tissue. Secondarily, it will be trained to assess the extent of lymphadenectomy and to differentiate between complete and incomplete lymph node dissection at each TIGER lymph node station. This will be done by semantic segmentation, the systematic practice of tagging or labeling data to train AI models, enabling them to process and interpret information in a manner similar to humans. Human-rated assessments (phase 3) will be used as a reference to train machine learning algorithms that allow for the automatic AI-based surgical quality assessment. This automatic quality assessment will exploit deep learning technology and is expected to perform an accurate and reproducible assessment of the photos/videos in our study. Subsequently, we will compare the performance of the resulting surgical AI to the performance of the human assessors. This final validation may also take the form of an international machine learning challenge, where computer scientists train their algorithms using our data, which we would then make openly available to the scientific community.

Data collection and statistical analysis

During the focus group discussions, a consensus agreement of 80% will be set as the benchmark for accepting the proposed elements of the structure of this study. To evaluate overall reliability, inter-rater reliability and inter-item reliability of the video and photo/video assessment tool, the generalizability theory (G-theory) will be applied. This is a statistical framework for conceptualizing, investigating, and designing reliable observations to determine the reliability (reproducibility) of measurements. The generalizability theory will be conducted using the G-String IV software.

To store and protect all data obtained and needed during this project, a secured digital platform will be used, developed in collaboration with FormsVision, a privately owned business company. This platform was previously developed for and has proven its functionality in the COLOR-III trial[16,19]. It allows for secured worldwide photo and video upload, as well as automatic email consignments to raters to assess the photo/video using the incorporated SQA tool. The database guarantees safety by an ALEA clinical processor agreement, and only essential people will have access to the data on the platform.

After completion of all assessments, the degree of surgical variability in the extent of lymphadenectomy will be quantified. We will adjust for any identified variability when analyzing the overall outcomes of the TIGER study [by using these data as explanatory variables (control measures) for study endpoints]. We additionally aim to clinically validate the SQA tool by comparing the traditional prognostic results from the TIGER study (such as lymph node yield and other postoperative data) against the results from the SQA tool (intraoperative data reflecting the quality of the lymphadenectomy), in their association with long-term survival and in their ability to be reflective of the prognostic value of lymphadenectomy. Logistic regressions on prognosis will be run for (i) lymph node yield and other postoperative parameters and (ii) intraoperative parameters. The comparison of the prognostic value will be based on the c-statistic. For all analyses, a *P*-value < 0.05 will be considered statistically significant. All analyses will be performed using IBM SPSS statistics.

DISCUSSION

The technically complex nature and the lack of widely agreed performance standards in surgery lead to variability in performances. These factors pose a significant challenge for RCTs investigating surgical innovations. When trials comparing two surgical techniques apply no methods of standardization, partial homogenization of the two study arms might appear, which will undermine the randomization[20]. In trials investigating the effect of oncologic interventions in addition to surgery, non-standardization of surgery might lead to different outcomes of the oncologic intervention for different groups.

Despite the above, RCTs are still considered the gold standard for surgical practices worldwide, and treatment protocols are mainly based on the results of these trials. To this end, SQA has been introduced in trials[16,21,22]. SQA includes either the upfront credentialing of surgeons or centers to participate in the trial, pretrial standardization of surgical techniques that are to be investigated in the trial, or monitoring performances during the trial[15,23]. Quality assurance programs in trials have been shown to reduce variation in outcome measures of surgical trials, such as lymph node yield, as highlighted by Markar *et al.* in their 2015 systematic review[15]. Future surgical trials should include an assessment of surgical performance as an important aspect of study design to reduce variation in clinical outcomes.

This SQA approach that incorporates credentialing, upfront standardization of surgical techniques, and monitoring to ensure consistent surgical performances across a study is not applicable to observational studies. In such studies, standards of surgery are not prescribed, and there is no intervention in the surgeons' performances. However, heterogeneity in the surgical performances probably affects the outcomes of observational studies and might compromise the generalizability of their results. Therefore, an alternative and more suitable SQA variant for observational studies (surgical quality assessment) is proposed to control the heterogeneity in surgical performances when analyzing outcome measures. However, it has not yet been performed.

In the international TIGER study, the distribution of lymph node metastases in esophageal carcinoma can only be reliably investigated if we accurately document not only how many and which lymph nodes are resected (as confirmed by the pathologist in the detailed pathology assessment), but also what remains unresected, i.e., if and how complete stations are cleared of lymphatic tissue by capturing the skeletonization of the structures left behind. Furthermore, studies examining the prognostic value of lymphadenectomy often rely on traditional pathological outcomes, such as lymph node yield, to quantify the extent of lymphadenectomy, as this is what is collected during surgery. However, providing an actual intraoperative metric - quantifying what happens in the operating room - may better reflect the quality of the lymphadenectomy, as opposed to relying on pathological proxies. Essentially, the hallmark of effective solid organ cancer surgery is ensuring that no malignant tissue is left behind. Therefore, it is imperative to develop a new metric to assess intraoperative surgical performance.

Emphasizing the importance of this TIGER-SQA study, we observed significant variation in the extent of lymphadenectomy in an international survey of 50 international expert upper gastrointestinal surgeons[14]. The number of different combinations of lymph node stations resected for the same tumor was notably high. Among the 14 surgeons who used the Japan Esophageal Society (JES) classification for proximal squamous cell cancer, none performed the same lymphadenectomy[24]. Another example of this problem is observed in the Dutch D1 D2 trial, which compared limited D1 with extended D2 lymphadenectomy during gastrectomy[20]. In this trial, 52% of the D1 operations were more extensive than the predefined D1 resection in the study protocol, while 84% of the D2 operations were less extensive than the pre-determined D2 dissection. This phenomenon entails that the chance of detecting a benefit for either of the two investigated techniques decreases, because both groups have become more homogeneous.

The integration of not only expert surgeon assessments but also AI-based automatic quality assessment systems for evaluating lymph node station removal has both strengths and limitations. One of the primary challenges associated with this approach is the potential for discrepancies between human and AI-based ratings. These inconsistencies may create uncertainty regarding which assessment should be considered the most accurate or reliable. While expert surgeon assessments offer clinical insights and practical expertise, they can be subject to variability due to human factors such as subjective interpretation of anatomical boundaries, completeness of lymphadenectomy, and fatigue. On the other hand, AI-driven assessments have the potential to improve consistency and reduce observer bias but require substantial development and validation. The AI-based system's ability to assess surgical quality relies heavily on the quality and diversity of the training data, which requires significant time and resources to collect. The complex and sometimes unpredictable nature of thoracic anatomy can present additional challenges, making it difficult for algorithms to perform accurately in rare or complicated cases. Therefore, it is anticipated that the development of a robust AI system capable of reliable performance across a wide range of scenarios may take years of research and iterative training. The current TIGER-SQA study, therefore, opts to use human assessments as the primary method due to these limitations. Nonetheless, future implementation of AI-based assessment in the evaluation of lymphadenectomy completeness during esophagectomy holds promise.

In conclusion, it is expected that in this TIGER-SQA study, the photographic and video clip assessment tool will be extended for both a 2- and 3-field lymphadenectomy in esophagectomy and that the developed SQA tool will be successfully used within the TIGER study. It will also lead to the development of an AI-based surgical assessment, which will make the assessment process more efficient, reproducible and achievable, considering the time constraints and inter- and intra-rater variation involved in having human surgeon assessors. This will be the first study to apply SQA for monitoring surgical performance within an observational study. It will ensure a reliable interpretation of TIGER study results and could make an important international contribution to our understanding of the prognostic value of the quality of surgical resection. Eventually, this study will provide a benchmark for SQA in future observational studies.

DECLARATIONS

Acknowledgments

The authors acknowledge the Dutch Cancer Society (KWF) for funding the current study (project number 14207), ALEA Clinical I FormsVision company (Abcoude, the Netherlands) for designing and developing the secured online database for this TIGER-SQA study, and the German Research Foundation for its funding of the current study (project ID 390696704).

Authors' contributions

Initiated the project and were major contributors in writing the protocol: Gisbertz SS, Henckens SPG
Contributed to the study design and writing the protocol on a day-to-day basis: Henckens SPG, Schuring N, Markar SR, van Berge Henegouwen MI, Hanna GB
Involved in the study design by attending several meetings in which the study was discussed: Harris A, Mavroveli S, Grüter AAJ, van der Aa DC, Wagner M, Daum M, Išgum I, Bodenstedt S, Speidel S, Müller-Stich BP
All authors read and approved the final manuscript.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Financial support and sponsorship

This TIGER-SQA study was funded by the Dutch Cancer Society (KWF, project number 14207), as well as the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy - EXC 2050/1 - Project ID 390696704 - Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI). KFT, DFG, and CeTI had no role in the design of the study, data collection, analysis, and interpretation, and writing of the manuscript.

Conflicts of interest

MüllerStich BP and Gisbertz SS are Editorial Board Members of the journal *Artificial Intelligence Surgery*. MüllerStich BP and Gisbertz SS were not involved in any steps of editorial processing, notably including reviewers' selection, manuscript handling and decision making. van Berge Henegouwen MI is a consultant for Johnson & Johnson, Alesi Surgical, BBraun, Mylan and Medtronic, and received research grants from Stryker. Išgum I received a research grant from Pie Medical Imaging, Philips Healthcare, Esaote, and is the co-founder of Quantib-U BV. The other authors declared that there are no conflicts of interest.

Ethics approval and consent to participate
The Medical Ethics Review Committee of the Academic Medical Center has confirmed that the Medical Research Involving Human Subjects Act (WMO) does not apply to the TIGER study because of the observational nature of the study. Written informed consent for prospective data collection was obtained from all patients prior to inclusion in the study. Official approval for this TIGER-SQA sub-study of the TIGER study by the IRB is therefore not required.

Consent for publication
Not applicable.

Copyright

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71:209-49. DOI PubMed
2. Lordick F, Mariette C, Haustermans K, Obermannová R, Arnold D; ESMO Guidelines Committee. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2016;27:v50-7. DOI PubMed
3. Kalff MC, Fransen LFC, de Groot EM, et al; Dutch Upper Gastrointestinal Cancer Audit group. Long-term survival after minimally invasive versus open esophagectomy for esophageal cancer: a nationwide propensity-score matched analysis. *Ann Surg* 2022;276:e749-57. DOI PubMed
4. Wang Y, Zhu L, Xia W, Wang F. Anatomy of lymphatic drainage of the esophagus and lymph node metastasis of thoracic esophageal cancer. *Cancer Manag Res.* 2018;10:6295-303. DOI PubMed PMC
5. Mariette C, Piessen G, Briez N, Triboulet JP. The number of metastatic lymph nodes and the ratio between metastatic and examined lymph nodes are independent prognostic factors in esophageal cancer regardless of neoadjuvant chemoradiation or lymphadenectomy extent. *Ann Surg.* 2008;247:365-71. DOI PubMed
6. Greenstein AJ, Litle VR, Swanson SJ, Divino CM, Packer S, Wisnivesky JP. Effect of the number of lymph nodes sampled on postoperative survival of lymph node-negative esophageal cancer. *Cancer.* 2008;112:1239-46. DOI PubMed
7. Castoro C, Scarpa M, Cagol M, et al. Nodal metastasis from locally advanced esophageal cancer: how neoadjuvant therapy modifies their frequency and distribution. *Ann Surg Oncol.* 2011;18:3743-54. DOI PubMed
8. Ruffato A, Mattioli S, Perrone O, et al. Esophagogastric metaplasia relates to nodal metastases in adenocarcinoma of esophagus and cardia. *Ann Thorac Surg.* 2013;95:1147-53. DOI PubMed
9. Tsurumaru M, Kajiyama Y, Udagawa H, Akiyama H. Outcomes of extended lymph node dissection for squamous cell carcinoma of the thoracic esophagus. *Ann Thorac Cardiovasc Surg.* 2001;7:325-9. PubMed
10. Li H, Zhang Y, Cai H, Xiang J. Pattern of lymph node metastases in patients with squamous cell carcinoma of the thoracic esophagus who underwent three-field lymphadenectomy. *Eur Surg Res.* 2007;39:1-6. DOI PubMed
11. Dresner SM, Lamb PJ, Bennett MK, Hayes N, Griffin SM. The pattern of metastatic lymph node dissemination from adenocarcinoma of the esophagogastric junction. *Surgery.* 2001;129:103-9. DOI PubMed
12. Pedrazzani C, de Manzoni G, Marrelli D, et al. Lymph node involvement in advanced gastroesophageal junction adenocarcinoma. *J Thorac Cardiovasc Surg.* 2007;134:378-85. DOI PubMed
13. Hagens E, Van Berge Henegouwen MI, Gisbertz S. PS02.158: distribution of lymph node metastases in esophageal carcinoma [TIGER study]: a multinational observational study. *Dis Esophagus.* 2018;31:166. DOI
14. van Rijswijk AS, Hagens ERC, van der Peet DL, van Berge Henegouwen MI, Gisbertz SS. Differences in esophageal cancer surgery in terms of surgical approach and extent of lymphadenectomy: findings of an international survey. *Ann Surg Oncol.* 2019;26:2063-72. DOI PubMed PMC
15. Markar SR, Wiggins T, Ni M, et al. Assessment of the quality of surgery within randomised controlled trials for the treatment of gastro-oesophageal cancer: a systematic review. *Lancet Oncol.* 2015;16:e23-31. DOI PubMed
16. Tsai AY, Mavroveli S, Miskovic D, et al. Surgical quality assurance in COLOR III: standardization and competency assessment in a randomized controlled trial. *Ann Surg.* 2019;270:768-74. DOI PubMed
17. Harris A, Butterworth J, Boshier PR, et al. Development of a reliable surgical quality assurance system for 2-stage esophagectomy in randomized controlled trials. *Ann Surg.* 2022;275:121-30. DOI PubMed
18. Schuring N, Matsuda S, Hagens ERC, et al. A proposal for uniformity in classification of lymph node stations in esophageal cancer. *Dis Esophagus.* 2021;34:doab009. DOI PubMed PMC
19. Deijen CL, Velthuis S, Tsai A, et al. COLOR III: a multicentre randomised clinical trial comparing transanal TME versus laparoscopic TME for mid and low rectal cancer. *Surg Endosc.* 2016;30:3210-5. DOI PubMed PMC
20. Bunt AM, Hermans J, Boon MC, et al. Evaluation of the extent of lymphadenectomy in a randomized trial of Western- versus

Japanese-type surgery in gastric cancer. *J Clin Oncol.* 1994;12:417-22. DOI PubMed

21. Han SU, Hur H, Lee HJ, et al. Surgeon quality control and standardization of D2 lymphadenectomy for gastric cancer: a prospective multicenter observational study (KLASS-02-QC). *Ann Surg.* 2021;273:315-24. DOI PubMed
22. Kim HI, Hur H, Kim YN, et al. Standardization of D2 lymphadenectomy and surgical quality control (KLASS-02-QC): a prospective, observational, multicenter study [NCT01283893]. *BMC Cancer.* 2014;14:209. DOI PubMed PMC
23. Foster JD, Mackenzie H, Nelson H, Hanna GB, Francis NK. Methods of quality assurance in multicenter trials in laparoscopic colorectal surgery: a systematic review. *Ann Surg.* 2014;260:220-9. DOI PubMed
24. Japan Esophageal Society. Japanese classification of esophageal cancer, 11th edition: part II and III. *Esophagus.* 2017;14:37-65. DOI PubMed PMC