

# Inter-observer and Intra-observer Variability in Gross Tumor Volume Delineation of Primary Esophageal Carcinomas Based on Different Combinations of Diagnostic Multimodal Images.

**Fengxiang Li**

Shandong Cancer Hospital and Institute, Shandong First Medical University; School of Medicine, Shandong University.

**Yankang Li**

School of Medicine, Shandong University; Shandong Cancer Hospital and Institute

**Xue Wang**

Linyi Cancer Hospital

**Jianbin Li** (✉ [lijianbin@msn.com](mailto:lijianbin@msn.com))

Shandong Cancer Hospital affiliated to Shandong University <https://orcid.org/0000-0002-4053-1792>

**Yingjie Zhang**

Shandong Cancer Hospital and Institute

**Xijun Liu**

Shandong Cancer Hospital and Institute

**Shanshan Liu**

Shandong Cancer Hospital and Institute

**Wei Wang**

Shandong Cancer Hospital and Institute

**Jinzhi Wang**

Shandong Cancer Hospital and Institute

**Yanluan Guo**

Shandong Cancer Hospital and Institute

**Min Xu**

Shandong Cancer Hospital and Institute

---

## Research

**Keywords:** Esophageal carcinoma, Diagnostic multimodal images, Target delineation, Intra-observer variability, Inter-observer variability

**Posted Date:** February 26th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-246391/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Frontiers in Oncology on April 1st, 2022. See the published version at <https://doi.org/10.3389/fonc.2022.817413>.

# Abstract

**Background and purpose** This study aimed to investigate inter-/ intra-observer delineation variability in GTVs of primary esophageal carcinomas (ECs) based on planning CT with reference to different combinations of diagnostic multimodal images from endoscopy/EUS, esophagography and FDG-PET/CT.

**Materials and methods** Fifty-one patients with pathologically proven thoracic EC who underwent diagnostic multimodal images before concurrent chemoradiotherapy were enrolled. Five radiotherapists independently delineated the GTVs based on planning CT only ( $GTV_C$ ), CT combined with endoscopy/EUS ( $GTV_{CE}$ ), CT combined with endoscopy/EUS and esophagography (X-ray) ( $GTV_{CEX}$ ), and CT combined with endoscopy/EUS, esophagography, and FDG-PET/CT ( $GTV_{CEXP}$ ). The intra-/inter-observer variability in the volume, longitudinal length, generalized CI ( $CI_{gen}$ ), and position of the GTVs were assessed.

**Results** The intra-/inter-observer variability in the volume and longitudinal length of the GTVs showed no significant differences ( $p>0.05$ ). The mean intra-observer  $CI_{gen}$  values for all observers was  $0.73\pm 0.15$ . The mean inter-observer  $CI_{gen}$  values for the four multimodal image combinations was  $0.67\pm 0.11$ . The inter-observer  $CI_{gen}$  for the four combined images was the largest, showing significant differences with those for the other three combinations. The intra-observer  $CI_{gen}$  among different observers and inter-observer  $CI_{gen}$  among different combinations of multimodal images showed significant differences ( $p<0.001$ ). The intra-observer  $CI_{gen}$  for the senior radiotherapists was larger than that for the junior radiotherapists ( $p<0.001$ ).

**Conclusion** For radiotherapists with advanced medical imaging training and clinical experience, using diagnostic multimodal images from endoscopy/EUS, esophagography, and FDG-PET/CT could reduce the intra-/inter-observer variability and increase the accuracy of target delineation in primary esophageal carcinomas.

## Highlights

- There is large variability in target volume delineation for esophageal carcinoma.
- Evaluation of Inter-/intra-observer delineation variability based on diagnostic multimodal imaging.
- Multimodal diagnostic image combinations can reduce the intra-/inter-observer variability and increase delineation accuracy.

## Introduction

Esophageal carcinoma is the seventh most commonly diagnosed cancer and the third leading cause of cancer deaths worldwide in 2018 [1]. Preoperative and definitive chemoradiation therapies have played a key role in the treatment of esophageal carcinoma [2–5]. The proportion of residual esophageal carcinoma after chemoradiation is significantly correlated with locoregional and distant failure [6–9]. Reliable delineation of the target gross tumor volume (GTV) is required for accurate radiation dose delivery and successful radiation therapy [10,11]. There is generally large variability in the target volume delineation for esophageal carcinoma, which might be primarily derived from the geometric uncertainties of different images and inherent variability among different observers based on the studies on other malignancies [12,13].

Conventional three-dimensional CT (3DCT) has been the workhorse modality used to delineate the esophageal tumor target volume. However, it is difficult to determine the proximal and distal extension of tumors and differentiate the layers of the esophageal wall [14–16]. An esophagography has shown a higher accuracy in assessing the tumor length (59% of cases, compared with 32% with CT), with tumor morphology influencing the accuracy [14–16]. Although endoscopy and endoscopic ultrasonography (EUS) might present the tumor length more accurately [17,18], it is difficult to transform the imaging to radiotherapy (RT) planning [19]. FDG-PET/CT has proved useful for diagnosing and staging esophageal carcinoma. However,

there is limited evidence supporting the validity of FDG-PET/CT for target volume delineation [20,21]. The false-positive FDG uptake in areas of inflammation reduces the specificity of tumor extent visualization [22]. Therefore, the combination of multimodal images is critical for determining the GTV of esophageal cancer (EC) accurately. Several studies have focused on the inter-observer variability of target volume delineation in FDG-PET/CT compared with pure CT imaging [10,23]. As CT imaging has proved indispensable for the visualization/detection of esophageal tumors, the use of multimodality imaging including esophagography, endoscopy/EUS and FDG-PET/CT for target volume delineation has not received sufficient attention.

In general, patients scheduled to receive radiotherapy or chemoradiotherapy undergo diagnostic multimodal imaging including enhanced CT, endoscopy/EUS, esophagography, or FDG-PET/CT. In clinical practice, radiation oncologists generally delineate the target volumes based on the planning CT images, with reference to various preexisting diagnostic images. However, the outcome of using different combinations of diagnostic multimodal images on the inter-observer and intra-observer delineation variability remains unclear. The purpose of this study was to investigate the inter-observer and intra-observer delineation variability in the GTVs of primary esophageal tumors with reference to different combinations of multimodal images from endoscopy/EUS, esophagography, and FDG-PET/CT.

## Materials And Methods

### *Patient selection and characteristics*

This study was approved by the ethics board of Shandong Cancer Hospital and Institute and informed consent has been obtained from the participants involved. Fifty-one patients with pathologically proven thoracic EC who had undergone preoperative or definitive concurrent chemoradiotherapy between May 2015 and June 2017 at Shandong Cancer Hospital and Institute were enrolled. Among the selected patients, there were seventeen cases each of upper, middle, and lower EC. One patient with lower EC was excluded due to the lack of PET-CT imaging data. All patients underwent a diagnostic imaging examination that included an endoscopy/EUS, esophagography, and FDG-PET/CT before receiving chemoradiotherapy. The average time for acquiring the diagnostic images was within the two-week period before chemoradiotherapy. Table 1 presents the patient characteristics.

### *Multimodal imaging*

**Endoscopy/EUS examination:** All patients underwent diagnostic endoscopy examinations using an electronic gastroscope (Olympus GIF-Q260J) before treatment. Seven patients did not undergo EUS examinations due to esophageal stenosis. The ultrasonic probe (Olympus EVIS EUS EU-ME2) was inserted into the patient's esophagus along the track of the biopsy forceps to detect the depth of tumor infiltration in the esophageal wall and the extent of proximal and distal tumor infiltration. The distances from the proximal and distal ends of the tumor to the incisors were recorded.

**Esophagography (X-ray) image acquisition:** Esophagography was performed before treatment using a digital radiography machine (Siemens Luminos dRT Max). All barium examinations were performed under fasting conditions, followed by a standard protocol (drinking 200 ml of diluted barium, in the upright, supine, and prone positions, with and without the gas powder).

**PET/CT image acquisition:** The PET-CT scan was performed within the two-week period prior to the planning CT scan as a part of the routine diagnostic management for EC. An 18F-FDG PET/CT scan of the chest was performed with an integrated PET/CT system (Philips Gemini TF Big Bore). The PET images were reconstructed with the CT-derived attenuation correction using an ordered subset expectation maximization algorithm with post-reconstruction Gaussian filtering, with a full width at half maximum of 5 mm.

**Planning CT image acquisition:** During the simulation, all patients were immobilized using a thermoplastic mask in the supine position with the arms placed along the side of the body. Each patient underwent an enhanced planning CT scan of the thoracic region on a 16-slice CT scanner (Philips Brilliance Bores CT) under free-breathing conditions. The planning CT images

were reconstructed using a thickness of 3 mm and subsequently transferred to an Eclipse treatment planning system (Varian Eclipse 11).

### *Target volume delineation*

A treatment planning system (Eclipse; Varian Medical Systems, Inc., Palo Alto, CA, USA) was used to contour the GTVs of the primary EC. The visualization parameter for delineation included the mediastinal window set to +40/400 HU. Before contouring, some clinical information such as the physical examination, pathological findings, and diagnostic CT image data were made available to the observers, while they were blind to the diagnostic endoscopy/EUS, esophagography, and FDG-PET/CT data. If the positive lymph nodes could not be separated from the primary tumor visually, they were delineated together with the primary tumor.

Five radiation oncologists (observers), who were blind to the diagnostic endoscopy/EUS, esophagography, and FDG-PET/CT patient data, were asked to independently delineate the GTVs with reference to different combinations of the multimodal images, including planning CT only (GTV<sub>C</sub>), CT combined with endoscopy/EUS (GTV<sub>CE</sub>), CT combined with endoscopy/EUS and esophagogram (X-ray) (GTV<sub>CEX</sub>), and CT combined with endoscopy/EUS, esophagogram, and FDG-PET/CT (GTV<sub>CEXP</sub>) (Figure 1). All observers were blind to the contours delineated by the other oncologists and their own former/previous contours. Observers 1 and 2 with clinical experience within five years were regarded as junior observers, while observers 3, 4, and 5 with more than ten years of clinical experience were regarded as senior observers. All contours were delineated in about two years. A delay of at least two months existed between each contouring of the tumor to eliminating a recall of the previous contouring for observers 1, 2, 3, and 5. The time interval for observer 4 was only one month, as the former observer 4 dropped out of the delineation process due to parturition.

### *Inter- / intra-observer variability analysis*

Inter-/intra-observer variability in the volume, longitudinal length, generalized conformity index (CI<sub>gen</sub>), and position of the GTVs was assessed. The intra-observer variability can be generally regarded as the variability of the same observer when re-contouring a single case. However, in this study, it is defined as the variability of the contours on the four multimodal imaging/image combinations for one observer [23].

The mean volume and longitudinal length of the GTVs based on different multimodal imaging combinations for different observers were calculated. The inter-observer variability in the volume and longitudinal length on different multimodal imaging combinations and the intra-observer variability for different observers were measured. The tumor length was measured using CT, endoscopy/EUS (43 cases), esophagography, and FDG-PET/CT, with the difference between the tumor length and corresponding longitudinal length of the GTVs subsequently evaluated.

CI<sub>gen</sub> was defined as the ratio of the common volume to encompassing volume [13,24]. The generalized CI (CI<sub>gen</sub>) was used to assess the overall consistency of all volume combinations delineated by different observers on the same imaging-modality combination, and that delineated by the same observer on different imaging-modality combinations. The formula is given by [13,25]:

$$CI_{gen} = \frac{\sum_{i,j} |A_i \cap A_j|}{\sum_{i,j} |A_i \cup A_j|}$$

CI<sub>gen</sub> is a good parameter for revealing the difference in the volumes delineated based on the size, shape, and location [10,23]. The use of CI<sub>gen</sub> tends to decrease the bias in the number of delineations [13]. The lower is the CI<sub>gen</sub> value for the same imaging-modality combination, the greater is the inter-observer variability. Similarly, a lower CI<sub>gen</sub> for the same observer suggests a greater intra-observer variability.

In addition, the x (right-left), y (anterior-posterior), and z (superior-inferior) axes of the center of mass (COM) of the volume were measured. The centroid shifts between the different volumes were then obtained. Finally, the three dimensional (3D) centroid shifts were calculated using the followed equation [24,26]:

$$3D \text{ centroid shifts} = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$$

### Statistical analysis

Statistical analysis was performed using the SPSS software package (SPSS 25.0). All the data had an approximately normal distribution. The one-way ANOVA test was applied to detect the inter-/intra-observer variability in the volume, longitudinal length,  $Cl_{gen}$ , and position of the GTVs among different observers and different multimodal imaging combinations. The paired *t*-test was used to compare the volume, longitudinal length,  $Cl_{gen}$ , and position of the GTVs between two observers or two multimodal imaging combinations. A  $P < 0.05$  was considered significant.

## Results

### GTV volume

Table 2 shows the primary GTV delineated based on four different multimodal imaging combinations for each observer. No significant inter-observer differences in the volume were observed for  $GTV_C$ ,  $GTV_{CE}$ ,  $GTV_{CEX}$ , or  $GTV_{CEXP}$  ( $p = 0.904, 0.987, 0.984, \text{ and } 0.97$ , respectively). The intra-observer variability in the volume of the GTVs derived from four different multimodal imaging combinations for observers 1–5 also showed no significant differences ( $p = 0.926, 0.997, 0.908, 0.943, \text{ and } 0.99$ , respectively). However, the paired comparisons indicated significant differences in the GTV volume between observers 1 and 2, observers 1 and 4, and observers 3 and 4 ( $t = 3.154, 6.368, \text{ and } 3.342, p = 0.002, <0.001, \text{ and } 0.001$ , respectively). Approximate statistical differences in the GTV volume were found between observers 1 and 3, and observers 2 and 4 ( $t = 3.342 \text{ and } 1.869, p = 0.061 \text{ and } 0.063$ , respectively).

### Esophageal tumor length

Table 3 shows the mean tumor lengths measured by CT, endoscopy/EUS, esophagography, and FDG-PET/CT. No significant differences were found between any two image-based tumor lengths. Table 3 presents the mean longitudinal lengths measured by the five observers corresponding to  $GTV_C$ ,  $GTV_{CE}$ ,  $GTV_{CEX}$ , and  $GTV_{CEXP}$ . The mean longitudinal length for  $GTV_{CEXP}$  was larger than the tumor length measured by FDG-PET/CT ( $p = 0.0035$ ). The intra-observer variability in the longitudinal length of the GTVs based on four multimodal imaging combinations for observers 1–5 showed no significant differences ( $p = 0.751, 0.794, 0.115, 0.962, \text{ and } 0.753$ , respectively). Table 2 shows the tumor lengths measured based on the four different multimodal imaging combinations for each observer. No significant inter-observer differences in the longitudinal length were recorded for  $GTV_C$ ,  $GTV_{CE}$ ,  $GTV_{CEX}$ , and  $GTV_{CEXP}$  ( $p = 0.286, 0.503, 0.997, \text{ and } 0.749$ , respectively). The two-related-samples tests indicated significant differences in the longitudinal lengths of the four GTVs between observers 1 and 2 ( $t = 2.776, p = 0.006$ ), observers 1 and 5 ( $t = 1.98, p = 0.049$ ), observers 3 and 2 ( $t = -3.166, p = 0.002$ ), and observers 3 and 5 ( $t = 2.992, p = 0.003$ ).

### Generalized CI ( $Cl_{gen}$ )

Table 4 lists the mean  $Cl_{gen}$  values for the four GTVs derived from different multimodal imaging combinations (mean intra-observer  $Cl_{gen}$ ) for each observer. The mean intra-observer  $Cl_{gen}$  values for all observers was  $0.73 \pm 0.15$ . The mean intra-observer  $Cl_{gen}$  was the largest for observer 4, exhibiting significant differences with that for the other observers. The mean intra-observer  $Cl_{gen}$  for observer 1 was the lowest, exhibiting significant differences with that for observers 3, 4, and 5. The mean intra-observer  $Cl_{gen}$  among different observers was statistically significant ( $F = 32.493, p < 0.001$ ). Table 5 lists the mean  $Cl_{gen}$  values for the five GTVs derived from different observers (mean inter-observer  $Cl_{gen}$ ) for each multimodal imaging

combination. The mean inter-observer  $Cl_{gen}$  values for the four multimodal imaging combinations was  $0.67 \pm 0.11$ . The mean inter-observer  $Cl_{gen}$  was the largest for the fourth multimodal imaging combination, which exhibited significant differences with that for the other three combinations. The mean inter-observer  $Cl_{gen}$  among the different multimodal imaging combinations showed a significant difference ( $F=6.872, p<0.001$ ).

### *Three-dimensional (3D) centroid shifts*

Table 4 lists the mean 3D centroid shifts of the four GTVs derived from different multimodal imaging combinations (mean intra-observer 3D centroid shifts) for each observer. The mean intra-observer 3D centroid shifts for all observers was  $3.67 \pm 4.62$  mm. The mean intra-observer 3D centroid shifts for observer 4 showed significant differences compared with the other observers. The mean intra-observer 3D centroid shifts among different observers was significant ( $F=3.898, p=0.004$ ). Table 5 presents the 3D centroid shifts of the five GTVs derived from different observers (mean inter-observer 3D centroid shifts) for each multimodal imaging combination. The mean inter-observer 3D centroid shifts for all four multimodal imaging combinations was  $3.81 \pm 4.7$  mm. The mean inter-observer 3D centroid shifts among the different multimodal imaging combinations showed no significant difference ( $F=0.327, p=0.806$ ).

## **Discussion**

Accurate target delineation for esophageal cancer is often restricted by the poor discriminative value of current imaging modalities [23], particularly CT, and the inability to relate diagnostic endoscopy/EUS, esophagography, or FDG-PET/CT information to the panning CT images [13–17,23]. However, reasonable pretreatment staging assessments are essential to determine a rational treatment strategy. In each patient with newly diagnosed esophageal cancer, the acquired diagnostic imaging information should identify the feasibility of delineating the GTVs of the primary based on the planning CT image with reference to the above-mentioned information. In this study, the geometric features of the GTVs derived from different observers and different planning CT image combinations were compared with the diagnostic imaging information. Furthermore, the value of the different planning CT image combinations in conjunction with diagnostic imaging information was evaluated for tumor delineation in esophageal carcinoma.

The results of this study showed no statistically significant inter-observer differences in the esophageal volume estimation based on different combinations of the CT, endoscopy/EUS, esophagography, and FDG-PET/CT data. For a particular multimodal imaging combination, different observers reported similar estimates for the GTV based on a similar knowledge of multimodal imaging. Moreover, for each observer, the volumes of the four GTVs delineated on different multimodal imaging combinations showed no significant differences. This indicates that the GTV volume assessments on different multimodal imaging combinations did not transform/change for the same observer. The data presented here is similar to the results reported in other literature [27,28]. However, Choi et al. [13] reported that the number of observers and number of observations made might affect the level of significance. In this study, many significant differences were observed in the GTV volume between different observers in the pairwise comparisons. Therefore, inter-observer variation in the target delineation could not be revealed/identified by merely comparing the volumes of the GTVs.

Similar to the observed variability in the volumes of the GTVs, the inter-observer and intra-observer variability in the longitudinal length showed no statistically significant differences. However, some significant differences between different observers were identified in the pairwise comparisons. The main reason behind these differences might be a different understanding of the procedure of determining the tumor length on multimodal imaging by different observers. Radiotherapists have always found the procedure to determine the proximal and distal extension of esophageal carcinoma based on different images challenging. Conventional images from CT, endoscopy/EUS, and esophagography, and MRI or FDG-PET/CT have their share of advantages and limitations for determining the tumor length [14–17,22,29,30]. It is critical to familiarize radiotherapists with these advantages and limitations before selecting the different image combinations. In this study, the tumor length determined by the multimodal images tended to be larger than that measured by a single image. In particular, the longitudinal length of  $GTV_{C_{EXP}}$  was significantly larger than the tumor length measured by FDG-PET/CT.

Therefore, the accuracy of tumor length measurements can be improved by combining the multimodal images; however, this is based on the precondition that the observers develop a good knowledge of the features of the multimodal images via unified training.

The  $Cl_{gen}$  values for  $GTV_C$ ,  $GTV_{CE}$ ,  $GTV_{CEX}$ , and  $GTV_{CEXP}$  for each observer represent the intra-observer variations, which include the random and inherent variations derived from different multimodal imaging combinations for the same observer. Here, the mean  $Cl_{gen}$  for intra-observer variability (0.73) was larger than that for inter-observer variability (0.67). This indicates that the intra-observer variability in delineating esophageal tumors was lower than the inter-observer variability, which shows agreement with the results reported in other studies [27,28]. Machiels et al. [27] reported the mean  $Cl_{gen}$  values for intra-observer delineation variability and inter-observer variability in ten patients without endoscopically implanted fiducial markers versus those with markers to be 0.54 versus 0.68 and 0.68 versus 0.75, respectively. Vollenbrock et al. [28] reported the mean  $Cl_{gen}$  over six patients as 0.68 on FDG-PET/CT, 0.66 on  $T_2w$ -MRI, and 0.68 on  $T_2w+DW$ (diffusion-weighted)-MRI. Compared with the above studies, fifty patients with upper, middle, and lower thoracic esophageal carcinoma were enrolled in this study. Moreover, different multimodal imaging combinations, including CT, endoscopy/EUS, esophagography, and FDG-PET/CT, were employed to eliminate any bias from a single imaging technique.

In the ANOVA analysis, the intra-observer  $Cl_{gen}$  for the GTVs derived from different multimodal imaging combinations among the five observers was statistically significant ( $p<0.001$ ). The intra-observer  $Cl_{gen}$  for the senior radiotherapists (observers 3, 4, and 5) was larger than that for the junior radiotherapists (observers 1 and 2). An optimum intra-observer  $Cl_{gen}$  was obtained for the senior radiotherapist who spent minimal time delineating the GTVs (observer 4). The senior radiotherapists, who were generally familiar with the multimodal imaging features for distinguishing the tumors from the normal structures and the location subject to relapse, might not be easily affected when only a single imaging modality is used/available for target contouring [31–32]. In addition, the shorter repeating delineation intervals did not eliminate the record of previous delineations, which might have improved the consistency of the target delineation. This suggests that background knowledge in medical imaging, clinical experience, and repeating delineation intervals might affect the intra-observer variability of the target  $Cl_{gen}$ .

While the inter-observer  $Cl_{gen}$  calculated for the different multimodal imaging combinations did not increase for the combined CT and endoscopy/EUS data, as compared with CT only,  $Cl_{gen}$  tended to increase for the combined CT and endoscopy/EUS and esophagography information ( $p=0.088$ ). Furthermore, the addition of FDG-PET/CT to the endoscopy/EUS, and esophagography data significantly improved the inter-observer  $Cl_{gen}$ . The use of multimodal imaging, including CT, endoscopy/EUS, esophagography, and FDG-PET/CT, for target delineation reduced the inter-observer variability. The effect of FDG-PET/CT on the intra- and inter-observer variability of target volume delineation in patients with gastro-esophageal cancer remains controversial. Vesprini et al. [33] reported that the combined use of FDG-PET/CT based on CT for GTV delineation significantly decreased both intra- and inter-observer variability, while Schreurs et al. [34] did not find PET/CT to have a significant effect on the inter-observer variability. Therefore, besides FDG-PET/CT, the additional use of endoscopy/EUS and esophagography for target delineation might prove beneficial in reducing the inter-observer variability. Recent studies have shown that the use of endoscopically implanted fiducial markers and MRI might reduce the variability of target volume delineation [27,28]. The use of multimodal imaging has proved increasingly valuable in improving the accuracy of target definition in esophageal carcinoma.

The variability in the position of the target volumes is one of the important factors influencing the  $Cl_{gen}$ . Here, the intra-observer variability in the 3D centroid shifts of the GTVs among different observers showed a significant difference. In the case of no significant variability in the volume of the GTVs, the intra-observer variability in the position could have mainly contributed to the statistical significance in the  $Cl_{gen}$  value. The intra-observer variability in the shape also tends to affect the intra-observer  $Cl_{gen}$ . The inter-observer variability in the 3D centroid shifts and volume of the GTVs showed no significant differences, suggesting that the inter-observer variability in the shape had a critical influence on the inter-observer  $Cl_{gen}$ . Thus, this study indirectly implies that using different multimodal image combinations might transform/change different observers' visual perception of tumors.

## Conclusion

In conclusion, for radiation oncologists with advanced medical imaging training and clinical experience, the use of diagnostic multimodal images from endoscopy/EUS, esophagography, and FDG-PET/CT for target delineation based on planning CT reduced the intra- and inter-observer variability and increased the accuracy of target delineation in primary thoracic esophageal carcinomas. In general, delineation of the GTV with reference to the diagnostic endoscopic and imaging data in primary EC is clearly useful. However, as this study did not identify a standard combination to determine the GTV, the effective implementation of diagnostic multimodal images for the delineation of primary esophageal carcinomas requires further clinical investigations.

## Abbreviations

**GTV** gross tumor volume

**3DCT** three-dimensional CT

**PET-CT** positron emission tomography-CT

**EUS** endoscopic ultrasonography

**EC** esophageal carcinomas

**Cigen** generalized conformity index

**COM** center of mass

## Declarations

- **Ethical Approval and Consent to participate**

This study was approved by the ethics board of Shandong Cancer Hospital and Institute. Informed consent has been obtained from the participants involved.

- **Consent for publication**

Not applicable.

- **Availability of supporting data**

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

- **Competing interests**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

- **Funding**

National Key Research Program of China (2016YFC0904700);

National Natural Science Foundation of China (817732870);

Key Research Program of Shandong Province (2016GSF201093);

- **Authors' contributions**

FXL and YKL equally contributed to the study design, the patient enrollment, the data statistics and analysis and writing the manuscript. JBL participated in the study design. XW contributed to the patient enrollment. YJZ, XJL, SSL, JZW, and MX contributed to the delineation. WW and YLG made important contributions in collecting the data and revising the content. All authors read and approved the final manuscript.

- **Acknowledgements**

Shandong Cancer Hospital and Institute

## References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
2. Sjoquist KM, Burmeister BH, Smithers BM, et al. Survival after neoadjuvant chemotherapy or chemoradiotherapy for resectable oesophageal carcinoma: an updated meta-analysis. *Lancet Oncol* 2011;12(7):681-92.
3. Shapiro J, van Lanschot JJB, Hulshof MCCM, et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *Lancet Oncol* 2015; 16(9):1090-8.
4. Lagergren J, Smyth E, Cunningham D, Lagergren P. Oesophageal cancer. 2017. 390(10110):2383-96.
5. Cooper JS, Guo MD, Herskovic A, et al. Chemoradiotherapy of locally advanced esophageal cancer: long-term follow-up of a prospective randomized trial (RTOG 85-01). *JAMA* 1999;281(17):1623-7.
6. Rohatgi PR, Swisher SG, Correa AM, et al. Failure patterns correlate with the proportion of residual carcinoma after preoperative chemoradiotherapy for carcinoma of the esophagus. *Cancer* 2005;104(7):1349-55.
7. Welsh J, Settle SH, Amini A, et al. Failure patterns in patients with esophageal cancer treated with definitive chemoradiation. *Cancer* 2012;118(10):2632-40.
8. Amini A, Ajani J, Komaki R, et al. Factors associated with local-regional failure after definitive chemoradiation for locally advanced esophageal cancer. *Ann Surg Oncol* 2014;21(1):306-14.
9. Xi M, Yang Y, Zhang L, et al. Multi-institutional Analysis of Recurrence and Survival After Neoadjuvant Chemoradiotherapy of Esophageal Cancer: Impact of Histology on Recurrence Patterns and Outcomes. *Ann Surg* 2019;269(4):663-70.
10. Nowee ME, Voncken FEM, Kotte ANTJ, et al. Dutch National Platform for Radiotherapy of Gastrointestinal Tumours (LPRGE) group. Gross tumour delineation on computed tomography and positron emission tomography-computed tomography in oesophageal cancer: A nationwide study. *Clin Transl Radiat Oncol* 2018;14:33-9.
11. Muijs C, Smit J, Karrenbeld A, et al. Residual tumor after neoadjuvant chemoradiation outside the radiation therapy target volume: a new prognostic factor for survival in esophageal cancer. *Int J Radiat Oncol Biol Phys* 2014; 88(4):845-52.
12. Steenbakkers RJ, Duppen JC, Fitton I, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis. *Int J Radiat Oncol Biol Phys* 2006;64(2):435-48.
13. Choi HJ, Kim YS, Lee SH, et al. Inter- and intra-observer variability in contouring of the prostate gland on planning computed tomography and cone beam computed tomography. *Acta Oncol* 2011;50(4):539-46.

14. Pollard JM, Wen Z, Sadagopan R, et al. The future of image-guided radiotherapy will be MR guided. *Br J Radiol* 2017;90:20160667.
15. Drudi FM, Trippa F, Cascone F, et al. Esophagogram and CT vs endoscopic and surgical specimens in the diagnosis of esophageal carcinoma. *Radiol Med* 2002; 103(4):344-52.
16. Encaoua J, Abgral R, Leleu C, et al. Radiotherapy volume delineation based on (<sup>18</sup>F)-fluorodeoxyglucose positron emission tomography for locally advanced or inoperable oesophageal cancer. *Cancer Radiother* 2017;21(4):267-75.
17. Caputo FM, Buquicchio GL. Esophageal cancer staging: the role of radiology. *Rays* 2005;30(4):309-14.
18. Wang B, Liu C, Lin C, et al. Endoscopic tumor length is an independent prognostic factor in esophageal squamous cell carcinoma. *Ann Surg Oncol*. 2012;19:2149–58.
19. Gao XS, Qiao X, Wu F, et al. Pathological analysis of clinical target volume margin for radiotherapy in patients with esophageal and gastroesophageal junction carcinoma. *Int J Radiat Oncol Biol Phys* 2007;67:389–96.
20. Muijs CT, Beukema JC, Pruim J, et al. A systematic review on the role of FDG-PET/CT in tumour delineation and radiotherapy planning in patients with esophageal cancer. *Radiother Oncol* 2010; 97(2):165-71.
21. Thomas L, Lapa C, Bundschuh RA, et al. Tumour delineation in oesophageal cancer – a prospective study of delineation in PET and CT with and without endoscopically placed clip markers. *Radiother Oncol* 2015;116:269–75.
22. Metser U, Even-Sapir E. Increased (<sup>18</sup>F)-fluorodeoxyglucose uptake in benign, nonphysiologic lesions found on whole-body positron emission tomography/computed tomography (PET/CT): accumulated data from four years of experience with PET/CT. *Semin Nucl Med* 2007;37:206–22.
23. Louie AV, Rodrigues G, Olsthorn J, et al. Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. *Radiother Oncol* 2010; 95(2):166-71.
24. Li FX, Li JB, Zhang YJ, et al. Geometrical differences in gross target volumes between 3DCT and 4DCT imaging in radiotherapy for non-small-cell lung cancer. *J Radiat Res* 2013; 54(5):950-6.
25. Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. *Phys Med Biol* 2009;54:2863 – 73.
26. Li FX, Li JB, Zhang YJ, et al. Comparison of the planning target volume based on three-dimensional CT and four-dimensional CT images of non-small-cell lung cancer. *Radiother Oncol* 2011 ;99(2):176-80.
27. Machiels M, Jin P, van Hooft JE, et al. Reduced inter-observer and intra-observer delineation variation in esophageal cancer radiotherapy by use of fiducial markers. *Acta Oncol* 2019; 58(6):943-50.
28. Vollenbrock SE, Nowee ME, Voncken FEM, et al. Gross Tumor Delineation in Esophageal Cancer on MRI Compared With 18F-FDG-PET/CT. *Adv Radiat Oncol* 2019 ;4(4):596-604.
29. Hou DL, Shi GF, Gao XS, et al. Improved longitudinal length accuracy of gross tumor volume delineation with diffusion weighted magnetic resonance imaging for esophageal squamous cell carcinoma. *Radiat Oncol* 2013; 6;8:169.
30. Machiels M, van Montfoort ML, Thuijs NB, et al. PO-0812: Pathological validation of endoscopically placed fiducials on tumor borders in esophageal cancer. *Radiother Oncol* 2019;133.
31. Van de Steene J, Linthout N, de Mey J, Vinh-Hung V, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. *Radiother Oncol* 2002 ;62(1):37-49.
32. Schimek-Jasch T, Troost EG, Rücker G, et al. A teaching intervention in a contouring dummy run improved target volume delineation in locally advanced non-small cell lung cancer: Reducing the interobserver variability in multicentre clinical studies. *Strahlenther Onkol* 2015 ;191(6):525-33.
33. Schreurs LM, Busz DM, Paardekooper GM, et al. Impact of 18- fluorodeoxyglucose positron emission tomography on computed tomography defined target volumes in radiation treatment planning of esophageal cancer: reduction in geographic misses with equal inter-observer variability. *Dis Esophagus* 2010;23(6):493-501.

34. Vesprini D, Ung Y, Dinniwell R, et al. Improving observer variability in target delineation for gastro-oesophageal cancer – the role of (18F)flfluoro-2-deoxy-Dglucose positron emission tomography/computed tomography. Clin Oncol (R Coll Radiol) 2008;20:631–8.

## Tables

Table 1

Patient's characteristics

| Characteristics                                   | Number               |
|---|----------------------|
| Sex, n(%)   |                      |
| M   | 40(80%)              |
| F   | 10(20%)              |
| Age, median, y(range)                             | 63(44-88)            |
| Tumor histology, n(%)                             |                      |
| Squamous cell carcinoma                           | 50(100%)             |
| SUV <sub>max</sub> , mean, median,y(range)        | 17.1, 15.2(2.8–49.5) |
| TNM* stage, n(%)                                  |                      |
| T <sub>2</sub> N <sub>0-2</sub> M <sub>0</sub>    | 4(8%)                |
| T <sub>3</sub> N <sub>1-3</sub> M <sub>0-1</sub>  | 34(68%)              |
| T <sub>4a</sub> N <sub>0-2</sub> M <sub>0-1</sub> | 12(24%)              |
| Tumor location, n(%)                              |                      |
| Upper   | 17(34%)              |
| Mid-  | 17(34%)              |
| Distal  | 16(32%)              |

Table 2

The volume and longitudinal length of GTVs based on different combinations of multimodal imaging for different observers

| Parameter                 |                          | Observer 1  | Observer 2  | Observer 3  | Observer 4  | Observer 5  | Mean±SD     |
|---------------------------|--------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| <b>GTV<sub>C</sub></b>    | Volume(cm <sup>3</sup> ) | 37.57±26.68 | 33.78±27.42 | 36.42±26.91 | 33.48±28.07 | 33.28±26.87 | 34.91±27.19 |
|                           | Length(mm)               | 5.7±2.6     | 5.3±2.5     | 6.3±2.5     | 5.7±2.6     | 5.3±2.3     | 5.7±2.5     |
| <b>GTV<sub>CE</sub></b>   | Volume(cm <sup>3</sup> ) | 34.06±25.90 | 34.47±26.65 | 36.33±27.10 | 32.68±25.79 | 35.13±27.50 | 34.68±26.82 |
|                           | Length(mm)               | 5.7±2.5     | 5.4±2.2     | 6.3±2.4     | 5.8±2.4     | 5.7±2.3     | 5.8±2.4     |
| <b>GTV<sub>CEX</sub></b>  | Volume(cm <sup>3</sup> ) | 35.43±25.79 | 35.00±26.40 | 33.57±26.68 | 34.50±26.84 | 35.03±28.37 | 34.34±26.81 |
|                           | Length(mm)               | 5.6±2.2     | 5.6±2.2     | 5.5±2.3     | 5.6±2.2     | 5.6±2.5     | 5.6±2.3     |
| <b>GTV<sub>CEXP</sub></b> | Volume(cm <sup>3</sup> ) | 36.30±27.41 | 34.17±26.19 | 33.48±27.13 | 33.28±26.67 | 36.73±28.27 | 35.04±27.17 |
|                           | Length(mm)               | 6.1±2.5     | 5.7±2.2     | 5.5±2.3     | 5.8±2.4     | 5.8±2.3     | 5.8±2.3     |
| <b>Mean±SD</b>            | Volume(cm <sup>3</sup> ) | 35.84±26.45 | 34.36±26.67 | 34.95±27.20 | 33.52±26.91 | 35.04±27.75 | 34.37±27.29 |
|                           | Length(mm)               | 5.8±2.4     | 5.5±2.3     | 5.9±2.4     | 5.7±2.4     | 5.6±2.4     | 5.7±2.4     |

Abbreviations: Gross target volumes (GTV) delineated on planning CT only (GTV<sub>C</sub>), CT combined with endoscopy/EUS (GTV<sub>CE</sub>), CT combined with endoscopy/EUS and esophagography (X-ray) (GTV<sub>CEX</sub>), and CT combined with endoscopy/EUS, esophagography, and FDG-PET/CT (GTV<sub>CEXP</sub>)

Table 3

Comparison the tumor length measured by CT, endoscopy/EUS, esophagography, and FDG-PET/CT with the mean longitudinal length measured by five observers for GTV<sub>C</sub>, GTV<sub>CE</sub>, GTV<sub>CEX</sub>, and GTV<sub>CEXP</sub>.

| Imaging modality               |                | CT                     | Endoscopy/EUS           | Esophagography           | PET-CT                    |
|--------------------------------|----------------|------------------------|-------------------------|--------------------------|---------------------------|
| <b>Tumor length(cm)</b>        |                | 5.5 ± 2.2              | 5.1 ± 2.0               | 5.3 ± 2.0                | 5.4 ± 2.2                 |
| <b>Target volume</b>           |                | <b>GTV<sub>C</sub></b> | <b>GTV<sub>CE</sub></b> | <b>GTV<sub>CEX</sub></b> | <b>GTV<sub>CEXP</sub></b> |
| <b>Longitudinal length(cm)</b> |                | 5.7 ± 2.5              | 5.8 ± 2.4               | 5.6 ± 2.3                | 5.8 ± 2.3                 |
| <b>Paired comparison</b>       | <b>t-value</b> | -0.704                 | -1.759                  | -1.272                   | -2.172                    |
|                                | <b>p-value</b> | 0.485                  | 0.086                   | 0.209                    | 0.035                     |

Abbreviations: Gross target volumes (GTV) delineated on planning CT only (GTV<sub>C</sub>), CT combined with endoscopy/EUS (GTV<sub>CE</sub>), CT combined with endoscopy/EUS and esophagography (X-ray) (GTV<sub>CEX</sub>), and CT combined with endoscopy/EUS, esophagography, and FDG-PET/CT (GTV<sub>CEXP</sub>)

Table 4

The  $CI_{gen}$  values and 3D centroid shifts (Mean $\pm$ SD) of the four GTVs derived from different combinations of multimodal imaging for each observer.

| Parameter            | Observer 1              | Observer 2              | Observer 3              | Observer 4              | Observer 5      |
|----------------------|-------------------------|-------------------------|-------------------------|-------------------------|-----------------|
| $CI_{gen}$           | 0.68 $\pm$ 0.12         | 0.69 $\pm$ 0.14         | 0.75 $\pm$ 0.15         | 0.80 $\pm$ 0.15         | 0.74 $\pm$ 0.17 |
| <b>Paired t-test</b> | 0.173(vs Obs2)          | -                       | -                       | -                       | -               |
| <b>(p value)</b>     | $\times$ 0.001(vs Obs3) | $\times$ 0.001(vs Obs3) | -                       | -                       | -               |
|                      | $\times$ 0.001(vs Obs4) | $\times$ 0.001(vs Obs4) | $\times$ 0.001(vs Obs4) | -                       | -               |
|                      | $\times$ 0.001(vs Obs5) | $\times$ 0.001(vs Obs5) | 0.264(vs Obs5)          | $\times$ 0.001(vs Obs5) | -               |
| <b>3D shifts(mm)</b> | 3.69 $\pm$ 4.47         | 4.34 $\pm$ 4.24         | 3.84 $\pm$ 3.93         | 2.85 $\pm$ 4.24         | 3.65 $\pm$ 6.19 |
| <b>Paired t-test</b> | 0.023(vs Obs2)          | -                       | -                       | -                       | -               |
| <b>(p value)</b>     | 0.572(vs Obs3)          | 0.077(vs Obs3)          | -                       | -                       | -               |
|                      | 0.005(vs Obs4)          | $\times$ 0.001(vs Obs4) | 0.001(vs Obs4)          | -                       | -               |
|                      | 0.911(vs Obs5)          | 0.084(vs Obs5)          | 0.59(vs Obs5)           | 0.039(vs Obs5)          | -               |

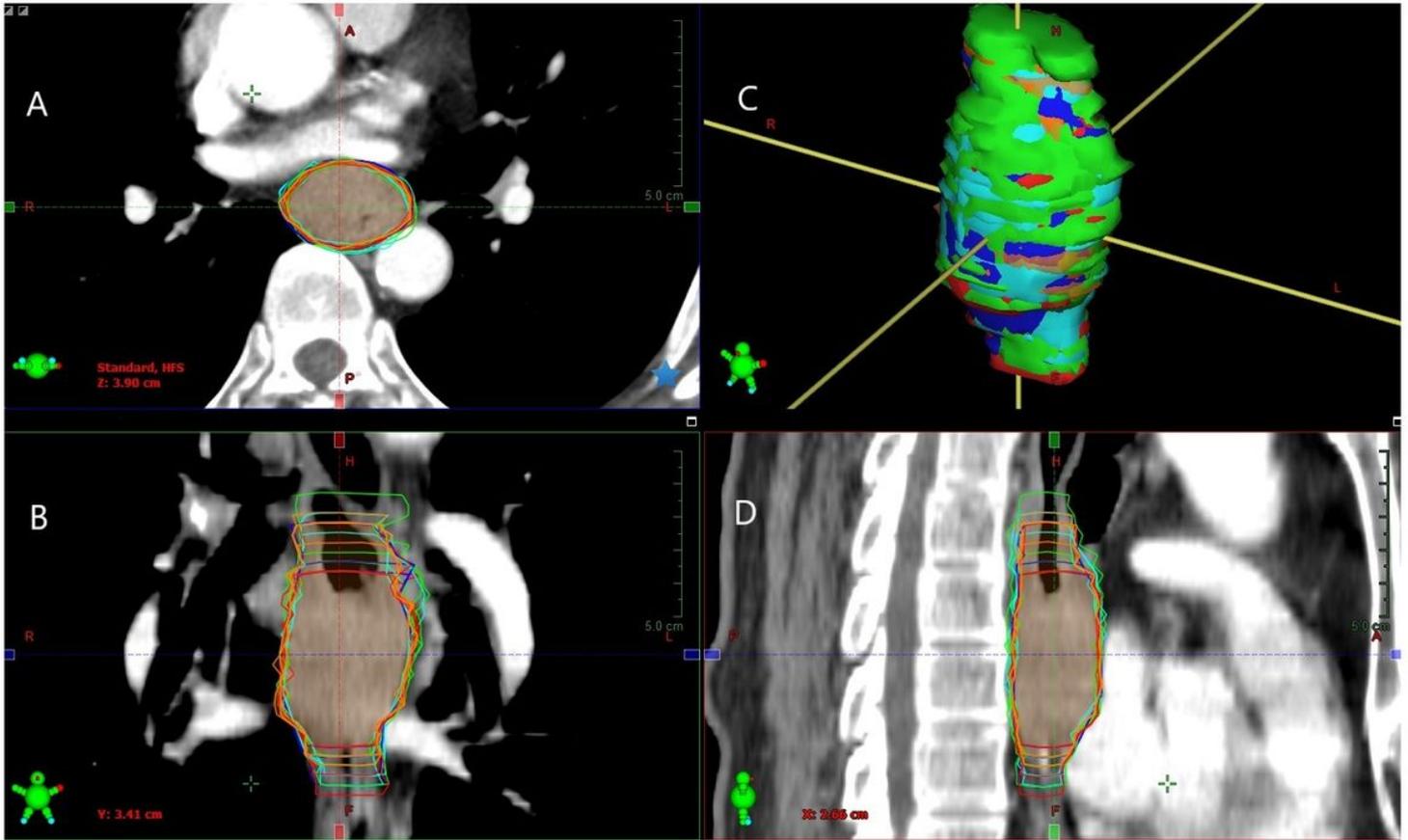
Table 5

The  $CI_{gen}$  values and 3D centroid shifts (Mean $\pm$ SD) of the five GTVs delineated by different observers based on each combinations of multimodal imaging.

| Parameter             | GTV <sub>C</sub>                      | GTV <sub>CE</sub>                     | GTV <sub>CEX</sub>                    | GTV <sub>CEXP</sub> |
|-----------------------|---------------------------------------|---------------------------------------|---------------------------------------|---------------------|
| $CI_{gen}$            | 0.66 $\pm$ 0.13                       | 0.66 $\pm$ 0.12                       | 0.67 $\pm$ 11                         | 0.69 $\pm$ 0.10     |
| <b>Paired t-test</b>  | 0.443(vs G <sub>CE</sub> )            | -                                     | -                                     | -                   |
| <b>(p value)</b>      | $\times$ 0.088(vs G <sub>CEX</sub> )  | 0.269(vs G <sub>CEX</sub> )           | -                                     | -                   |
|                       | $\times$ 0.001(vs G <sub>CEXP</sub> ) | $\times$ 0.001(vs G <sub>CEXP</sub> ) | $\times$ 0.001(vs G <sub>CEXP</sub> ) | -                   |
| <b>3D shifts (mm)</b> | 3.78 $\pm$ 4.04                       | 3.78 $\pm$ 3.79                       | 3.98 $\pm$ 5.03                       | 3.68 $\pm$ 5.94     |
| <b>Paired t-test</b>  | 0.981(vs G <sub>CE</sub> )            | -                                     | -                                     | -                   |
| <b>(p value)</b>      | 0.463(vs G <sub>CEX</sub> )           | 0.463(vs G <sub>CEX</sub> )           | -                                     | -                   |
|                       | 0.762(vs G <sub>CEXP</sub> )          | 0.744(vs G <sub>CEXP</sub> )          | 0.218(vs G <sub>CEXP</sub> )          | -                   |

Abbreviations: Gross target volumes (GTV) delineated on planning CT only (GTV<sub>C</sub>), CT combined with endoscopy/EUS (GTV<sub>CE</sub>), CT combined with endoscopy/EUS and esophagography (X-ray) (GTV<sub>CEX</sub>), and CT combined with endoscopy/EUS, esophagography, and FDG-PET/CT (GTV<sub>CEXP</sub>)

## Figures



**Figure 1**

Example of GTVs delineated based on different combinations of multimodal images by observer 1 (green segment), observer 2 (red segment), observer 3 (blue segment), observer 4 (orange segment), and observer 5 (cyan segment) in tansversal (A), frontal (B), surface (C), and sagittal (D) planes for one patient (Patient 5).