

E. P. M. van Vliet  
J. J. Hermans  
W. De Wever  
M. J. C. Eijkemans  
E. W. Steyerberg  
C. Faasse  
E. P. M. van Helmond  
A. M. de Leeuw  
A. C. Sikkenk  
A. R. de Vries  
E. H. de Vries  
E. J. Kuipers  
P. D. Siersema

## Radiologist experience and CT examination quality determine metastasis detection in patients with esophageal or gastric cardia cancer

Received: 19 February 2008

Accepted: 19 April 2008

Published online: 4 June 2008

© The Author(s) 2008

E. P. M. van Vliet (✉) · E. J. Kuipers ·  
P. D. Siersema  
Department of Gastroenterology and  
Hepatology, Erasmus MC–University  
Medical Center Rotterdam,  
P.O. Box 2040, 3000  
CA Rotterdam, The Netherlands  
e-mail: evelynvanvliet@hotmail.com  
Tel.: +31-10-4634681  
Fax: +31-10-4634682

J. J. Hermans  
Department of Radiology, Erasmus  
MC–University Medical Center  
Rotterdam,  
P.O. Box 2040, 3000  
CA Rotterdam, The Netherlands

W. De Wever  
Department of Radiology,  
University Hospitals Gasthuisberg,  
Leuven, Belgium

M. J. C. Eijkemans · E. W. Steyerberg  
Department of Public Health, Erasmus  
MC–University Medical Center  
Rotterdam,  
P.O. Box 2040, 3000  
CA Rotterdam, The Netherlands

C. Faasse  
Department of Radiology,  
Franciscus Hospital,  
Roosendaal, The Netherlands

E. P. M. van Helmond  
Department of Radiology,  
Harbour Hospital,  
Rotterdam, The Netherlands

A. M. de Leeuw  
Department of Radiology,  
Beatrix Hospital,  
Gorinchem, The Netherlands

A. C. Sikkenk  
Department of Radiology,  
Medical Centre Rijnmond-South,  
Rotterdam, The Netherlands

A. R. de Vries  
Department of Radiology,  
Albert Schweitzer Hospital,  
Dordrecht, The Netherlands

E. H. de Vries  
Department of Radiology,  
Vlietland Hospital,  
Vlaardingen, The Netherlands

### Present address:

P. D. Siersema  
Department of Gastroenterology and  
Hepatology, University Medical Center  
Utrecht,  
Utrecht, The Netherlands

**Abstract** We aimed to separate the influence of radiologist experience from that of CT quality in the evaluation of CT examinations of patients with esophageal or gastric cardia cancer. Two radiologists from referral centers ('expert radiologists')

and six radiologists from regional non-referral centers ('non-expert radiologists') performed 240 evaluations of 72 CT examinations of patients diagnosed with esophageal or gastric cardia cancer between 1994 and 2003. We used conditional logistic regression analysis to calculate odds ratios (OR) for the likelihood of a correct diagnosis. Expert radiologists made a correct diagnosis of the presence or absence of distant metastases according to the gold standard almost three times more frequently (OR 2.9; 95% CI 1.4–6.3) than non-expert radiologists. For the subgroup of CT examinations showing distant metastases, a statistically significant correlation (OR 3.5; 95% CI 1.4–9.1) was found between CT quality as judged by the radiologists and a correct diagnosis. Both radiologist experience and quality of the CT examination play a role in the detection of distant metastases in esophageal or gastric cardia cancer patients. Therefore, we suggest that staging procedures for esophageal and gastric cardia cancer should preferably be performed in centers with technically advanced equipment and experienced radiologists.

**Keywords** Staging · CT · Esophageal cancer · Gastric cardia cancer

## Introduction

Staging of patients with esophageal or gastric cardia cancer is of utmost importance for the selection of an optimal treatment, which should be aggressive in cases with potential of cure, but more limited when only palliation of symptoms is possible. As a result of staging investigations, the extent of local invasion of the tumor through the esophageal wall is represented by the T stage, the presence of regional lymph node metastases by the N stage, and the presence of distant metastases by the M stage [1]. Esophageal cancer most commonly metastasizes to regional, celiac, and supraclavicular lymph nodes; as well as liver, lung, and adrenal glands [2].

In the Netherlands, upper gastrointestinal (GI) endoscopies are mostly performed in regional non-referral centers, i.e., centers without specific expertise in the treatment of upper GI cancer. Following a diagnosis of esophageal or gastric cardia cancer, preoperative staging investigations are often performed in these regional centers. Thereafter, patients may be treated in these centers or are referred to a specialized referral center. In the latter situation, staging investigations performed in the referring regional center are usually re-evaluated and/or repeated in the referral center [3].

In a previous retrospective study, we found that metastases in patients with esophageal or gastric cardia cancer were more frequently detected on computed tomography (CT) examinations made and evaluated in a referral center compared with regional centers. We speculated that this was caused by the presence of more experienced radiologists and/or the use of technically more advanced equipment in the referral center [3]. In the current study, we aimed to separate radiologist experience from CT quality to determine the exact role of these factors in the evaluation of CT examinations of patients with esophageal or gastric cardia cancer.

## Materials and methods

### Patients

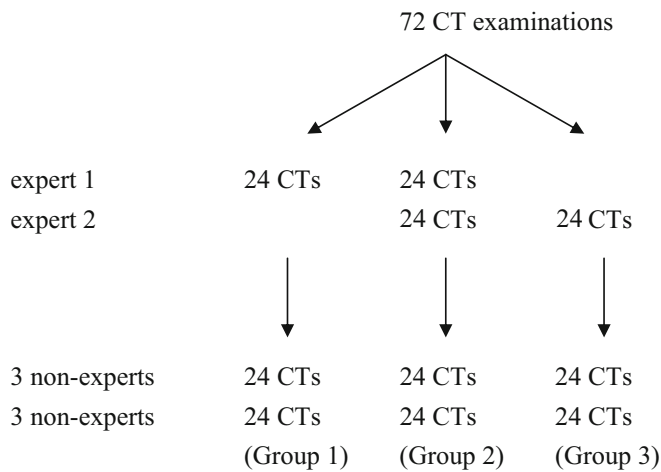
In the Erasmus MC–University Medical Center Rotterdam, the Netherlands, a database is maintained with information on 1,088 patients who have been diagnosed with esophageal or gastric cardia cancer in the period January 1994 to October 2003. In 906/1,088 patients, the diagnosis was made in a regional center and, subsequently, these patients were referred to our referral center for further evaluation and/or treatment. In 477/906 patients, no CT examination was performed in the regional center ( $n=340$ ) or the CT performed in the regional center was not re-evaluated and/or repeated in the referral center ( $n=137$ ). In the remaining 429 patients, the CT examination performed in the regional center was re-evaluated and/or repeated. In 235/429

patients, the CT examination was only re-evaluated in our referral center (re-evaluated CTs), which occurred when the quality of the CT examination was determined as sufficient. In 79/429 patients, the quality of the CT examination performed in the regional center was determined as insufficient after re-evaluation, and the CT was therefore subsequently repeated (re-evaluated and repeated CTs). In 115/429 patients, the CT examination was not re-evaluated in the referral center, but was immediately repeated, which became particularly common after 1999 when a new generation of CT systems became available in the referral center (repeated CTs).

### Methods

Two radiologists from two different referral centers for esophageal and gastric cardia cancer ('expert radiologists') and six radiologists from six different regional centers ('non-expert radiologists') evaluated hard copy CT examinations of patients with esophageal or gastric cardia cancer. The two referral centers had a volume of more than 100 patients with esophageal or gastric cardia cancer per year, whereas the total number of patients with esophageal or gastric cardia cancer in regional centers was low, i.e., less than ten cases per year. As a result, radiologists from the referral centers, with 8 and 13 years of experience respectively, evaluated more CTs of patients with this malignancy on a yearly basis than radiologists from the regional centers, making them 'expert radiologists' in the current study. All radiologists knew that they were part of a study and considered to be an expert or non-expert radiologist.

We made a selection of 72 hard copy CT examinations out of all CTs performed in the 429 patients with a previously re-evaluated and/or repeated CT. This selection was stratified according to the following schedule: 26 re-evaluated CTs from regional centers, 28 repeated CTs (14 from regional centers and 14 from the referral center performed in the same 14 patients), and 18 re-evaluated and repeated CTs (nine from regional centers and nine from the referral center performed in the same nine patients). This was an almost similar distribution as compared with the number of re-evaluated and/or repeated CTs in the period January 1994 to October 2003 in our center. We decided to include CTs from both regional centers and the referral center that were performed in the same patient, as possible differences found could then only be attributed to the origin of the CTs and not to the characteristics of the patients. Although the number of patients with distant metastases in our center is lower than the number of patients without distant metastases, we decided to choose for a distribution in which the number of CTs with distant metastases was more or less equal to those without distant metastases. CT examinations were randomly selected taking into account the criteria mentioned above. Of all



**Fig. 1** Distribution of the CT examinations among the various radiologists

72 CTs evaluated, 37 (51%) had distant metastases, whereas the other 35 (49%) were without. Celiac lymph node metastases were considered as regional (N1) if the primary tumor was located in the gastric cardia and as distant metastases (M1) if the tumor was located in the esophagus. The gold standard was the postoperative pathological TNM stage, the result of fine-needle aspiration (FNA), or a radiological result with  $\geq 6$  months of follow-up. In patients with the gold standard postoperative pathological TNM stage or the result of FNA, no new metastases were found in the 6 months following resection or FNA, which suggests that the results of these gold standards were reliable. None of the patients received neoadjuvant therapy that could have changed the disease status.

The distribution of CT examinations among the participating radiologists is shown in Fig. 1. We made three groups of 24 CT examinations. The two expert radiologists evaluated 48 CTs, of which one set of 24 CT examinations was evaluated by both expert radiologists to determine the variability between these radiologists. The six non-expert radiologists each evaluated 24 CTs. In order to determine the variability between non-expert radiologists, the 24 CTs in each group were evaluated by two non-expert radiologists. In summary, in group 1 and 3, 24 CTs were evaluated by one expert and two non-expert radiologists. In group 2, 24 CTs were evaluated by two expert and two non-expert radiologists.

Each radiologist evaluated CT examinations from regional centers and the referral center (Table 1). In addition, each radiologist randomly evaluated two different CTs of the same patient, meaning that the CT from the regional center and that from the referral center were evaluated by the same radiologist. In group 1, four CTs from the regional center and four CTs from the referral center performed in the same patients were evaluated by the radiologists. In group 2, this number was 6 and in group 3, it was 5. Furthermore, each radiologist evaluated CTs with metastases as well as CTs without metastases according to the gold standard (Table 1).

CT examination quality was determined with four criteria, which were given a score: (a) whether or not intravenous contrast medium was administered (bolus enhanced) (yes, 1; no, 2), (b) slice thickness ( $\leq 8$  mm, 1;  $> 8$  mm, 2), (c) completeness of the CT examination (neck/thorax/abdomen, 1; missing part, 2), and (d) whether or not lung window settings were included (yes, 1; no, 2). This resulted in a score varying between 4 and 8, in which CT examinations with a score of 4 were considered to be of

**Table 1** Characteristics of CT examinations per group of 24 CTs

	Group 1 <i>n</i> =24	Group 2 <i>n</i> =24	Group 3 <i>n</i> =24	Total <i>n</i> =72
Origin of CT examination				
Regional center	16	17	16	49
Referral center	8	7	8	23
Number of patients with CT examination from the regional center and from the referral center	4	6	5	23
N stage				
N0	17	11	13	41
N1	7	13	11	31
M stage				
M0	12	11	12	35
M1	12	13	12	37
Gold standard				
Radiological finding with $\geq 6$ months of follow-up	9	8	5	22
Fine-needle aspiration	6	4	8	18
Postoperative stage	9	12	11	32

good quality and those with a score of 8 were of poor quality.

The radiologists evaluated the CTs using a standardized form on: (a) quality (good/moderate/poor/too poor to evaluate), (b) presence of tumor (yes/no), and, if 'yes', the primary location (esophagus/gastroesophageal junction/gastric cardia), and (c) presence of metastases (yes/no), and, if 'yes', the location. No objective measures for CT quality were given to the radiologists. All radiologists used normal daily clinical life criteria which they considered to be indicated for a CT examination of good, moderate, or poor quality and, therefore, this grading of quality was subjective.

The study has been carried out with ethical committee approval.

### Statistical analysis

The results of the evaluations of the radiologists were compared with the gold standard, i.e., the postoperative pathological TNM stage, the result of fine-needle aspiration (FNA), or a radiological result with  $\geq 6$  months of follow-up. Sensitivities, specificities, and accuracies for N and M stage were calculated per radiologist. In addition, sensitivities and specificities were calculated for metastases per organ, i.e., metastases in regional and celiac lymph nodes, lung, liver, and adrenal glands. For each group, results of the radiologists were compared to determine whether radiologist experience was important in the evaluation of CT examinations. The McNemar test was performed to determine whether the differences between the results of N and M stage were statistically significant. To determine whether the hospital where the CT had been performed was important, sensitivities and specificities for N and M stage were calculated for CTs from regional centers and from the referral center per radiologist. Furthermore, multivariable conditional logistic regression analysis was performed to determine the relative importance of radiologist experience and CT origin, adjusting for the statistical clustering of multiple CTs of the same patient.

To determine whether CT quality was important, sensitivities and specificities for N and M stage were calculated for CTs scored as good or moderate by the radiologists and for CTs scored as poor or too poor to evaluate. These results were calculated per radiologist. In addition, multivariable conditional logistic regression analysis was repeated with CT quality according to the opinion of each radiologist as an extra covariate in the model.

In addition, conditional logistic regression analysis was performed for the subgroup of CTs without lymph node or distant metastases (specificity) and CTs with lymph node or distant metastases according to the gold standard (sensitivity), to determine whether the role of radiologist experience and origin and quality of CT examination were the same for the detection as well as the exclusion of metastatic disease.

Chi-square testing was used to determine whether a correlation was present between CT quality according to the opinion of radiologists on the one hand and the quality score, radiologist experience, and CT origin on the other hand.

Software used for the analyses were SPSS (SPSS version 12.0, Chicago, IL) and EGRET (EGRET version 2, Cytel Software Corporation, Cambridge, MA). All  $p$  values were based on two-sided tests. A  $p$  value  $< 0.05$  was considered as statistically significant.

## Results

### Per radiologist

#### *Radiologist experience*

In Table 2, sensitivities, specificities, and accuracies are shown for each radiologist. The results for N stage differed per group. For example, in group 1, the highest sensitivity for N stage was obtained by expert 1 with lower sensitivities for non-experts 1 and 2, whereas in group 2 the highest sensitivity was obtained by expert 2 and non-expert 3 with lower sensitivities for expert 1 and non-expert 4. Accuracies for M stage were slightly higher for expert than for non-expert radiologists. These differences were however not statistically significant. Five of the eight radiologists had not evaluated all CT examinations as they judged the quality of some CTs too poor to allow evaluation. For that reason, we also calculated sensitivities, specificities, and accuracies of CT examinations that were evaluated by all radiologists in each separate group and these results were compared with the data shown in Table 2. We found that the results of CTs that were evaluated by all radiologists of each group (data not shown) were higher than the results shown in Table 2.

#### *Origin of CT examination*

No correlations were found between CT findings and the hospital where the CT examination had been performed (Table 3). However, a correlation was found between the hospital where the CT had been performed and the quality according to the radiologist. Radiologists gave significantly higher quality scores to CT examinations from the referral center than to those from the regional centers (Table 4).

#### *Quality of CT examination*

In addition to the CT origin, we also looked at the correlation between quality scores given by the radiologists and CT findings. Sensitivities for N and M stage were

**Table 2** Results of evaluated CT examinations per radiologist

	Group 1			Group 2				Group 3		
	Expert 1 (n=24)	Non-expert 1 (n=24)	Non-expert 2 (n=22)	Expert 1 (n=21)	Expert 2 (n=20)	Non-expert 3 (n=22)	Non-expert 4 (n=23)	Expert 2 (n=22)	Non-expert 5 (n=24)	Non-expert 6 (n=24)
<b>Sensitivity (%)</b>										
N stage	4/7 (57)	4/7 (57)	1/6 (17)	4/13 (31)	6/11 (55)	9/13 (69)	4/13 (31)	3/9 (33)	4/11 (36)	6/11 (55)
M stage	7/12 (58)	3/12 (25)	1/11 (9)	6/12 (50)	3/10 (30)	9/13 (69)	3/12 (25)	6/10 (60)	4/12 (33)	4/12 (33)
Regional lymph nodes	4/7 (57)	4/7 (57)	1/6 (17)	4/13 (31)	6/11 (55)	9/13 (69)	4/13 (31)	3/9 (33)	4/11 (36)	6/11 (55)
Celiac lymph nodes	7/10 (70)	1/10 (10)	0/9 (0)	3/6 (50)	0/4 (0)	2/7 (29)	0/6 (0)	0/5 (0)	0/7 (0)	0/7 (0)
Lung	0/1 (0)	0/1 (0)	0/1 (0)	1/2 (50)	0/2 (0)	1/2 (50)	0/2 (0)	–	–	–
Liver	0/2 (0)	1/2 (50)	1/2 (50)	2/4 (50)	3/4 (75)	2/4 (50)	1/3 (25)	4/5 (80)	4/5 (80)	2/5 (40)
Adrenal gland	–	–	–	–	–	–	–	–	–	–
<b>Specificity (%)</b>										
N stage	14/17 (82)	9/17 (53)	16/16 (100)	4/8 (50)	8/9 (89)	5/9 (56)	7/10 (70)	10/13 (77)	11/13 (85)	10/13 (77)
M stage	10/12 (83)	9/12 (75)	10/11 (91)	9/9 (100)	10/10 (100)	5/9 (56)	11/11 (100)	12/12 (100)	9/12 (75)	10/12 (83)
Regional lymph nodes	14/17 (82)	9/17 (53)	16/16 (100)	4/8 (50)	8/9 (89)	5/9 (56)	7/10 (70)	10/13 (77)	11/13 (85)	10/13 (77)
Celiac lymph nodes	14/14 (100)	14/14 (100)	13/13 (100)	15/15 (100)	16/16 (100)	10/15 (67)	17/17 (100)	17/17 (100)	16/17 (94)	14/17 (82)
Lung	23/23 (100)	22/23 (96)	21/21 (100)	18/19 (95)	18/18 (100)	17/20 (85)	21/21 (100)	22/22 (100)	19/24 (79)	24/24 (100)
Liver	20/22 (91)	19/22 (86)	19/20 (95)	17/17 (100)	16/16 (100)	15/18 (83)	19/19 (100)	15/17 (88)	17/19 (90)	18/19 (95)
Adrenal gland	24/24 (100)	24/24 (100)	22/22 (100)	19/21 (91)	20/20 (100)	21/22 (96)	21/23 (91)	22/22 (100)	24/24 (100)	24/24 (100)
<b>Accuracy (%)</b>										
N stage	18/24 (75)	13/24 (54)	17/22 (77)	8/21 (38)	14/20 (70)	14/22 (64)	11/23 (49)	13/22 (59)	15/24 (63)	16/24 (67)
M stage	17/24 (71)	12/24 (50)	11/22 (50)	15/21 (71)	13/20 (65)	14/22 (64)	14/23 (61)	18/22 (82)	13/24 (54)	14/24 (58)

Each group of 24 CTs was evaluated by expert and non-expert radiologists. Five of the eight radiologists had not evaluated all CT examinations as they judged the quality of some CTs too poor to allow evaluation

higher for CT examinations of good or moderate quality compared with those of poor quality, whereas specificities for N and M stage were lower (Table 5).

Nine of 72 CTs (12%) had a quality score of 4 (see [Methods](#)), 20 CTs (28%) of 5, 15 CTs (21%) of 6, 20 CTs (28%) of 7, and eight CTs (11%) of 8. A correlation was found between CT quality according to the radiologists and the calculated quality score ( $p < 0.001$ ). CTs with a score of 4 or 5 were more often judged as of good quality, whereas CTs with a score of 7 or 8 were often considered too poor to evaluate. There was no correlation between CT quality

as judged by the radiologists and radiologist experience ( $p = 0.18$ ).

#### Conditional logistic regression analyses

##### *Radiologist experience*

Conditional logistic regression analysis for N stage showed no statistically significant correlation between radiologist experience (expert/non-expert) and a correct diagnosis of

**Table 3** Sensitivity and specificity of CT examinations from the regional center and the referral center

	Group 1			Group 2				Group 3		
	Expert 1 (n=24)	Non-expert 1 (n=24)	Non-expert 2 (n=22)	Expert 1 (n=21)	Expert 2 (n=20)	Non-expert 3 (n=22)	Non-expert 4 (n=23)	Expert 2 (n=22)	Non-expert 5 (n=24)	Non-expert 6 (n=24)
Regional center	n=16	n=16	n=14	n=14	n=13	n=15	n=16	n=14	n=16	n=16
Sensitivity (%)										
N stage	2/4 (50)	2/4 (50)	1/3 (33)	3/9 (33)	4/7 (57)	6/9 (67)	3/9 (33)	2/7 (29)	2/9 (22)	5/9 (56)
M stage	3/7 (43)	2/7 (29)	1/6 (17)	2/8 (25)	1/6 (17)	5/9 (56)	2/8 (25)	4/7 (57)	3/9 (33)	2/9 (22)
Specificity (%)										
N stage	11/12 (92)	5/12 (42)	11/11 (100)	2/5 (40)	5/6 (83)	4/6 (67)	5/7 (71)	6/7 (86)	5/7 (71)	5/7 (71)
M stage	8/9 (89)	7/9 (78)	8/8 (100)	6/6 (100)	7/7 (100)	4/6 (67)	8/8 (100)	7/7 (100)	6/7 (86)	7/7 (100)
Referral center	n=8	n=8	n=8	n=7	n=7	n=7	n=7	n=8	n=8	n=8
Sensitivity (%)										
N stage	2/3 (67)	2/3 (67)	0/3 (0)	1/4 (25)	2/4 (50)	3/4 (75)	1/4 (25)	1/2 (50)	2/2 (100)	1/2 (50)
M stage	4/5 (80)	1/5 (20)	0/5 (0)	4/4 (100)	2/4 (50)	4/4 (100)	1/4 (25)	2/3 (67)	1/3 (33)	2/3 (67)
Specificity (%)										
N stage	3/5 (60)	4/5 (80)	5/5 (100)	2/3 (67)	3/3 (100)	1/3 (33)	2/3 (67)	4/6 (67)	6/6 (100)	5/6 (83)
M stage	2/3 (67)	2/3 (67)	2/3 (67)	3/3 (100)	3/3 (100)	1/3 (33)	3/3 (100)	5/5 (100)	3/5 (60)	3/5 (60)

Radiologists from the same group evaluated same CTs. Five of the eight radiologists had not evaluated all CT examinations as they judged the quality of some CTs too poor to allow evaluation

the presence or absence of lymph node metastases according to the gold standard (odds ratio (OR) 0.9; 95% confidence interval (CI) 0.5–1.8) (Table 6). A correlation was also not found for the subgroup of CT examinations without lymph node metastases (specificity) and CTs with lymph node metastases (sensitivity), indicating that radiologist experience does not assist in determining N stage.

Conditional logistic regression analysis for M stage showed that expert radiologists almost three times more frequently made a correct diagnosis of the presence or absence of distant metastases than non-expert radiologists (OR 2.9; 95% CI 1.4–6.3) (Table 6). For the subgroup of CT examinations without distant metastases this chance was nearly seven times higher (OR 6.9; 95% CI 1.3–37.0). The association was less pronounced for the subgroup of CTs with distant metastases (OR 2.2; 95% CI 0.9–5.5). These results indicate that radiologist experience is

important in determining M stage, particularly for confirming the absence of distant metastases.

#### Origin of CT examination

Conditional logistic regression analysis for N and M stage showed no statistically significant correlation between CT origin and a correct diagnosis (Table 6), indicating that the origin of a CT is not important for detecting metastases in patients with esophageal or gastric cardia cancer.

#### Quality of CT examination

Both for N and M stage, conditional logistic regression analysis showed no correlation between the radiologists'

**Table 4** Correlation between the CT origin and the quality according to the opinion of the radiologists

Origin	Quality			
	Good	Moderate	Poor	Too poor to evaluate
Referral center	41 (59%)	26 (30%)	8 (13%)	0 (0%)
Regional center	28 (41%)	60 (70%)	52 (87%)	20 (100%)

Linear-by-linear association test:  $p < 0.001$

**Table 5** Sensitivity and specificity of CT examinations judged as of good or moderate quality or of poor quality according to the opinion of each radiologist

	Group 1			Group 2				Group 3		
	Expert 1 (n=24)	Non-expert 1 (n=24)	Non-expert 2 (n=22)	Expert 1 (n=21)	Expert 2 (n=20)	Non-expert 3 (n=22)	Non-expert 4 (n=23) <sup>a</sup>	Expert 2 (n=22)	Non-expert 5 (n=24) <sup>b</sup>	Non-expert 6 (n=24)
Quality score: good or moderate	n=16	n=17	n=16	n=14	n=17	n=9	n=15	n=14	n=14	n=22
Sensitivity (%)										
N stage	4/6 (67)	4/5 (80)	0/5 (0)	3/8 (38)	6/9 (67)	4/5 (80)	4/10 (40)	3/6 (50)	2/3 (67)	6/10 (60)
M stage	6/8 (75)	2/10 (20)	0/9 (0)	3/7 (43)	3/9 (33)	5/6 (83)	3/9 (33)	4/7 (57)	4/8 (50)	4/11 (36)
Specificity (%)										
N stage	8/10 (80)	6/12 (50)	11/11 (100)	4/6 (67)	7/8 (88)	2/4 (50)	4/5 (80)	7/8 (88)	9/11 (82)	9/12 (75)
M stage	7/8 (88)	5/7 (71)	6/7 (86)	7/7 (100)	8/8 (100)	1/3 (33)	6/6 (100)	7/7 (100)	3/6 (50)	9/11 (82)
Quality score: poor/ too poor to evaluate	n=8	n=7	n=6	n=7	n=3	n=13	n=7	n=8	n=6	n=2
Sensitivity (%)										
N stage	0/1 (0)	0/2 (0)	1/1 (100)	1/5 (20)	0/2 (0)	5/8 (63)	0/2 (0)	0/3 (0)	0/4 (0)	0/1 (0)
M stage	1/4 (25)	1/2 (50)	1/2 (50)	3/5 (60)	0/1 (0)	4/7 (57)	0/2 (0)	2/3 (67)	0/4 (0)	0/1 (0)
Specificity (%)										
N stage	6/7 (86)	3/5 (60)	5/5 (100)	0/2 (0)	1/1 (100)	3/5 (60)	3/5 (60)	3/5 (60)	2/2 (100)	1/1 (100)
M stage	3/4 (75)	4/5 (80)	4/4 (100)	2/2 (100)	2/2 (100)	4/6 (67)	5/5 (100)	5/5 (100)	2/2 (100)	1/1 (100)

Radiologists from the same group evaluated the same CTs. Five of the eight radiologists had not evaluated all CT examinations as they judged the quality of some CTs too poor to allow evaluation

<sup>a</sup>One CT examination without information on quality

<sup>b</sup>Four CT examinations without information on quality

opinion on CT quality and a correct diagnosis (OR 0.9; 95% CI 0.6–1.6 and OR 1.9; 95% CI 1.0–3.7, respectively) (Table 6). For the subgroup of CT examinations with distant metastases, it was however found that the chance to confirm the presence of distant metastases was 3.5 times higher for a 1-point-higher quality score compared with a lower quality score as judged by a radiologist. This indicates that CT quality is an important factor in confirming the presence of distant metastases.

## Discussion

In this study, two expert and six non-expert radiologists performed 240 evaluations of 72 hard copy CT examinations of patients diagnosed with esophageal or gastric cardia cancer to determine whether radiologist experience and/or CT quality were factors involved in the evaluation of CT examinations. Our findings showed that expert radiologists more frequently made a correct diagnosis of the presence or absence of distant metastases than non-expert radiologists. For the subgroup of CTs with distant

metastases, a correlation was found between the radiologists' opinion on CT quality and a correct diagnosis, which indicates that, in addition to expertise, CT quality also plays a role in detecting distant metastases from esophageal or gastric cardia cancer. However, radiologist experience and CT quality did not play a role in determining N stage.

Accuracies, sensitivities, and specificities for N and M stage differed remarkably between radiologists in the present study. This variation for N stage evaluation in esophageal cancer has also been reported in the literature, i.e., from 33 to 86% for accuracy, 22 to 84% for sensitivity, and 60 to 100% for specificity [4–27]. The same is also true for M stage, with reported accuracies, sensitivities, and specificities in the ranges 45–94%, 32–81%, and 11–97%, respectively [4, 6, 8, 17, 22, 24, 25, 27–30]. The differences between different radiologists found in our study were however statistically not significant (Table 2). This could be due to various reasons. First, the number of evaluated CT examinations in this study might have been too low to detect statistically significant differences between radiologists. Alternatively, it could be that differences between expert and non-expert radiologists were indeed small, which makes

**Table 6** Conditional logistic regression analyses analyzing whether a correlation is present between radiologist experience, CT quality and CT origin on the one hand and a correct diagnosis according to the gold standard on the other hand

Variable	Odds ratio (95% CI)
Radiologist experience (expert versus non-expert) <sup>a</sup>	
Lymph node metastases	
All CT examinations	0.94 (0.50–1.77)
CT examinations with metastases according to gold standard	0.68 (0.25–1.86)
CT examinations without metastases according to gold standard	1.23 (0.52–2.89)
Distant metastases	
All CT examinations	2.93 (1.36–6.29)
CT examinations with metastases according to gold standard	2.21 (0.89–5.52)
CT examinations without metastases according to gold standard	6.90 (1.29–37.0)
Origin of CT examination (referral center versus regional center) <sup>a</sup>	
Lymph node metastases	
All CT examinations	1.06 (0.46–2.42)
CT examinations with metastases according to gold standard	1.31 (0.37–4.62)
CT examinations without metastases according to gold standard	1.22 (0.52–2.89)
Distant metastases	
All CT examinations	0.85 (0.38–1.94)
CT examinations with metastases according to gold standard	0.46 (0.16–1.35)
CT examinations without metastases according to gold standard	2.38 (0.56–10.09)
Quality of CT examination <sup>b</sup>	
Lymph node metastases	
All CT examinations	0.93 (0.56–1.55)
CT examinations with metastases according to gold standard	0.91 (0.41–2.03)
CT examinations without metastases according to gold standard	1.04 (0.53–2.05)
Distant metastases	
All CT examinations	1.94 (1.00–3.68)
CT examinations with metastases according to gold standard	3.52 (1.36–9.08)
CT examinations without metastases according to gold standard	0.78 (0.28–2.17)

CI confidence interval

<sup>a</sup>Covariates are radiologist experience and origin of CT examination<sup>b</sup>Covariates are radiologist experience, origin of CT examination, and quality of CT examination (continuous, range 1=good to 4=too poor to evaluate)

them clinically irrelevant. The first reason seems most likely, as we found a correlation between radiologist experience and a correct diagnosis of the presence or absence of distant metastases in the conditional logistic regression analysis (Table 6). The confidence intervals for the ORs are wide, which might also be due to the relatively low number of CTs that were evaluated.

We think that our finding that expert radiologists were more likely to make a correct diagnosis with regard to the presence or absence of distant metastases than non-expert radiologists is due to differences in radiologist experience. It may, however, also be due to differences in evaluation practices between expert and non-expert radiologists. For example, it may be that expert radiologists are less inclined to report the presence of distant metastases than non-expert radiologists. This will lead to fewer false-positive results

(higher specificity), but also to more false-negative results (lower sensitivity) for expert radiologists. Nevertheless, the opposite may also be true, i.e., that expert radiologists more frequently have false-positive results (lower specificity), but fewer false-negative results (higher sensitivity) than non-expert radiologists. Nonetheless, this study suggests that obvious differences in evaluation practices were not present. For the subgroup of CTs without distant metastases (specificity), and to a lesser extent for those with distant metastases (sensitivity), the OR for radiologist experience was above 1 (Table 6), which indicates that expert radiologists were more likely to make a correct diagnosis than non-expert radiologists for both subgroups of CT examinations.

We also determined whether CT quality was important for the detection of lymph node and distant metastases. For



this, the radiologists were asked to give an opinion on CT quality. Objective measures for CT quality were not given to the radiologists and all radiologists used normal daily clinical life criteria which they considered to be indicated for a CT examination of good, moderate, or poor quality, where they took into account, among other factors, use of contrast medium, slice thickness, and completeness of the CT examination. Remarkably, the number of CT examinations of good or moderate quality ranged from 9 to 22 (Table 5), which shows that ranking CT quality is a subjective matter, in which some radiologists were more decisive to give lower quality scores than other radiologists.

The chance to confirm the presence of distant metastases was higher for higher quality than for lower quality CT examinations (OR 3.5; 95% CI 1.4–9.1) (Table 6). This suggests that distant metastases in patients with esophageal or gastric cardia cancer are not always visible or not easily detected on CTs of poor quality. A correlation was not found for the subgroup of CTs without distant metastases according to the gold standard (OR 0.8; 95% CI 0.3–2.2) (Table 6), which indicates that CT quality is less important in confirming the absence of distant metastases. This finding is clinically important as patients with distant metastases should preferably undergo a palliative treatment and not a surgical resection [31].

We also analyzed whether CT origin (regional/referral center) correlated with a correct diagnosis of the presence or absence of lymph node or distant metastases. In the referral center, the newest-generation CT systems were used during the study period, which was not always true for the regional centers. In addition, intravenous and oral contrast medium was always administered in the referral center, whereas CTs without contrast medium were performed in some of the regional centers. Particularly, liver metastases are more readily detected after contrast enhancement [32, 33]. Our analyses showed no correlation between CT origin and a correct diagnosis (Table 6). Nevertheless, a correlation was found between CT origin and quality according to the opinion of the radiologists, with higher quality scores for CTs from the referral center (Table 4).

There are some limitations to this study. First, the study was not performed in daily clinical practice. The radiologists who evaluated CT examinations were all aware of the fact that these CTs had been made in patients with esophageal or gastric cardia cancer, but had no information on the results of other staging investigations. In clinical practice, radiologists are however not always blinded to these results. Furthermore, 51% of the evaluated CT examinations were performed in patients with distant metastases according to the gold standard, whereas the other CTs were of patients without distant metastases. This distribution is not like daily clinical practice, with fewer

patients having distant metastases. However, we chose this distribution, as the number of CT examinations with distant metastases would otherwise have been too small to draw conclusions. In addition, CT examinations will often be reviewed by a multidisciplinary team in daily clinical practice, which was not the case in our study, in which radiologists evaluated the CTs alone.

Second, not all radiologists evaluated all available CT examinations, as they judged the quality of some CTs too poor to allow a conclusion to be made. We proposed that CTs that were not evaluated were specifically those for which it was more difficult to determine whether or not lymph node or distant metastases were present. To determine whether this was indeed the case, we also calculated sensitivities, specificities, and accuracies of CT examinations that were evaluated by all radiologists in each separate group. We found that the results of CTs that were evaluated by all radiologists of each group were higher than the results shown in Table 2. This suggests that CT examinations that were not evaluated by radiologists were indeed CTs for which it was more difficult to determine whether lymph node or distant metastases were present.

Finally, this study was based on hard copy image data sets, as the CT examinations used in this study were not digitally available. Furthermore, some CT examinations were incomplete, meaning that not the complete thorax and/or abdomen was present on the examination. Particularly, the lung and liver could not be fully evaluated in some cases. There are two possible explanations for these incomplete CTs. First, slides might have been lost over the years. Second, the ‘missing’ slides were not made due to the protocol that was used in that center. We assume that the latter explanation was most likely, as the ‘missing’ slides were always slides above the highest or below the lowest part of the body that was investigated.

In conclusion, both radiologist experience and quality of CT examination are important factors in the evaluation of CT examinations performed in patients with esophageal or gastric cardia cancer. The results from this study suggest that staging procedures for esophageal or gastric cardia cancer should preferably be performed in centers with the ability to produce high quality CTs and by radiologists with ample experience in evaluating CT examinations for this indication, which will optimally allow the detection of distant metastases from esophageal or gastric cardia cancer.

**Acknowledgments** The first author of this article was funded by a grant from the ‘Doelmatigheidsonderzoek’ fund of the Erasmus MC–University Medical Center Rotterdam, the Netherlands.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Fleming ID, Cooper JS, Henson DE (1997) AJCC cancer staging manual. Lippincott-Raven, Philadelphia
- Quint LE, Hepburn LM, Francis IR et al (1995) Incidence and distribution of distant metastases from newly diagnosed esophageal carcinoma. *Cancer* 76:1120–1125
- van Vliet EP, Eijkemans MJ, Kuipers EJ et al (2006) A comparison between low-volume referring regional centers and a high-volume referral center in quality of preoperative metastasis detection in esophageal carcinoma. *Am J Gastroenterol* 101:234–242
- Lea JWt, Prager RL, Bender HW Jr (1984) The questionable role of computed tomography in preoperative staging of esophageal cancer. *Ann Thorac Surg* 38:479–481
- Quint LE, Glazer GM, Orringer MB, Gross BH (1985) Esophageal carcinoma: CT findings. *Radiology* 155:171–175
- Tio TL, Cohen P, Coene PP et al (1989) Endosonography and computed tomography of esophageal carcinoma. Preoperative classification compared to the new (1987) TNM system. *Gastroenterology* 96:1478–1486
- Ziegler K, Sanft C, Zeitz M et al (1991) Evaluation of endosonography in TN staging of oesophageal cancer. *Gut* 32:16–20
- Botet JF, Lightdale CJ, Zauber AG et al (1991) Preoperative staging of esophageal cancer: comparison of endoscopic US and dynamic CT. *Radiology* 181:419–425
- Sondenaa K, Skaane P, Nygaard K, Skjennald A (1992) Value of computed tomography in preoperative evaluation of resectability and staging in oesophageal carcinoma. *Eur J Surg* 158:537–540
- Yoshikane H, Tsukamoto Y, Niwa Y et al (1994) Superficial esophageal carcinoma: evaluation by endoscopic ultrasonography. *Am J Gastroenterol* 89:702–707
- Greenberg J, Durkin M, Van Drunen M, Aranha GV (1994) Computed tomography or endoscopic ultrasonography in preoperative staging of gastric and esophageal tumors. *Surgery* 116:696–701
- Holden A, Mendelson R, Edmunds S (1996) Pre-operative staging of gastro-oesophageal junction carcinoma: comparison of endoscopic ultrasound and computed tomography. *Australas Radiol* 40:206–212
- Flanagan FL, Dehdashti F, Siegel BA et al (1997) Staging of esophageal cancer with 18F-fluorodeoxyglucose positron emission tomography. *AJR Am J Roentgenol* 168:417–424
- Massari M, Cioffi U, De Simone M et al (1997) Endoscopic ultrasonography for preoperative staging of esophageal carcinoma. *Surg Laparosc Endosc* 7:162–165
- Nishimaki T, Tanaka O, Ando N et al (1999) Evaluation of the accuracy of preoperative staging in thoracic esophageal cancer. *Ann Thorac Surg* 68:2059–2064
- Choi JY, Lee KH, Shim YM et al (2000) Improved detection of individual nodal involvement in squamous cell carcinoma of the esophagus by FDG PET. *J Nucl Med* 41:808–815
- Wren SM, Stijns P, Srinivas S (2002) Positron emission tomography in the initial staging of esophageal cancer. *Arch Surg* 137:1001–1007
- Kienle M, Buhl K, Kuntz C et al (2002) Prospective comparison of endoscopy, endosonography and computed tomography for staging of tumours of the oesophagus and gastric cardia. *Digestion* 66:230–236
- Drudi FM, Trippa F, Cascone F et al (2002) Esophagogram and CT vs endoscopic and surgical specimens in the diagnosis of esophageal carcinoma. *Radiol Med (Torino)* 103:344–352
- Vazquez-Sequeiros E, Wiersema MJ, Clain JE et al (2003) Impact of lymph node staging on therapy of esophageal carcinoma. *Gastroenterology* 125:1626–1635
- Wu LF, Wang BZ, Feng JL et al (2003) Preoperative TN staging of esophageal cancer: comparison of miniprobe ultrasonography, spiral CT and MRI. *World J Gastroenterol* 9:219–224
- Rasanen JV, Sihvo EI, Knuuti MJ et al (2003) Prospective analysis of accuracy of positron emission tomography, computed tomography, and endoscopic ultrasonography in staging of adenocarcinoma of the esophagus and the esophagogastric junction. *Ann Surg Oncol* 10:954–960
- Yoon YC, Lee KS, Shim YM et al (2003) Metastasis to regional lymph nodes in patients with esophageal squamous cell carcinoma: CT versus FDG PET for presurgical detection prospective study. *Radiology* 227:764–770
- Kneist W, Schreckenberger M, Bartenstein P et al (2003) Positron emission tomography for staging esophageal cancer: does it lead to a different therapeutic approach? *World J Surg* 27:1105–1112
- Sihvo EI, Rasanen JV, Knuuti MJ et al (2004) Adenocarcinoma of the esophagus and the esophagogastric junction: positron emission tomography improves staging and prediction of survival in distant but not in locoregional disease. *J Gastrointest Surg* 8:988–996
- Heeren PA, Jager PL, Bongaerts F et al (2004) Detection of distant metastases in esophageal cancer with (18)F-FDG PET. *J Nucl Med* 45:980–987
- Lowe VJ, Booya F, Fletcher JG et al (2005) Comparison of positron emission tomography, computed tomography, and endoscopic ultrasound in the initial staging of patients with esophageal cancer. *Mol Imaging Biol* 7:422–430
- Luketich JD, Friedman DM, Weigel TL et al (1999) Evaluation of distant metastases in esophageal cancer: 100 consecutive positron emission tomography scans. *Ann Thorac Surg* 68:1133–1137
- Flamen P, Lerut A, Van Cutsem E et al (2000) Utility of positron emission tomography for the staging of patients with potentially operable esophageal carcinoma. *J Clin Oncol* 18:3202–3210
- Kneist W, Schreckenberger M, Bartenstein P et al (2004) Prospective evaluation of positron emission tomography in the preoperative staging of esophageal carcinoma. *Arch Surg* 139:1043–1049
- Siersema PD (2006) New developments in palliative therapy. *Best Pract Res Clin Gastroenterol* 20:959–978
- Salonen O, Kivisaari L, Standertskjold-Nordenstam CG et al (1987) Computed tomography in staging of oesophageal carcinoma. *Scand J Gastroenterol* 22:65–68
- Kirk SJ, Moorehead RJ, McIlrath E et al (1990) Does preoperative computed tomography scanning aid assessment of oesophageal carcinoma? *Postgrad Med J* 66:191–194