Original article

Pragmatic staging of oesophageal cancer using decision theory involving selective endoscopic ultrasonography, PET and laparoscopy

J. M. Findlay[1,2], K. M. Bradley[3], E. J. Maile[1], B. Braden[4], J. Maw[1], J. Phillips-Hughes[3], R. S. Gillies[1], N. D. Maynard[1] and M. R. Middleton[2,5]

[1]Oxford OesophagoGastric Centre, [2]National Institute for Health Research Oxford Biomedical Research Centre and [3]Department of Radiology, Churchill Hospital and [4]Department of Gastroenterology, John Radcliffe Hospital, Oxford University Hospitals NHS Trust, and [5]Department of Oncology, University of Oxford, Oxford, UK

Correspondence to: Mr J. M. Findlay, Oxford OesophagoGastric Centre, Churchill Hospital, Old Road, Oxford OX3 7LJ, UK
(e-mail: john.findlay@oncology.ox.ac.uk)

Background: **Following CT, guidelines for staging oesophageal and gastro-oesophageal junction (GOJ) cancer recommend endoscopic ultrasonography (EUS), PET–CT and laparoscopy for T3–T4 GOJ tumours. These recommendations are based on generic utilities, but it is unclear whether the test risk outweighs the potential benefit for some patients. This study sought to quantify investigation risks, benefits and utilities, in order to develop pragmatic, personalized staging recommendations.**

Methods: **All patients with a histological diagnosis of oesophageal or GOJ cancer staged between May 2006 and July 2013 comprised a development set; those staged from July 2013 to July 2014 formed the prospective validation set. Probability thresholds of altering management were calculated and predictive factors identified. Algorithms and models (decision tree analysis, logistic regression, artificial neural networks) were validated internally and independently.**

Results: **Some 953 patients were staged following CT, by [^{18}F]fluorodeoxyglucose PET–CT (918), EUS (798) and laparoscopy (458). Of these patients, 829 comprised the development set (800 PET–CT, 698 EUS, 397 laparoscopy) and 124 the validation set (118 PET–CT, 100 EUS, 61 laparoscopy). EUS utility in the 71·8 per cent of patients with T2–T4a disease on CT was minimal (0·4 per cent), its risk exceeding benefit. EUS was moderately accurate for pT1 N0 disease. A number of factors predicted metastases on PET–CT and laparoscopy, although none could inform an algorithm. PET–CT altered management in 23·0 per cent, and laparoscopy in 7·1 per cent, including those with T2 and distal oesophageal tumours.**

Conclusion: **Although EUS provided additional information on T and N category, its risk outweighed potential benefit in patients with T2–T4a disease on CT. Laparoscopy seemed justified for distal oesophageal tumours of T2 or greater.**

Presented in part to meetings of the Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland, Newcastle upon Tyne, UK, September 2013, the Society of Academic and Research Surgery, Cambridge, UK, January 2014, and the Association of Surgeons of Great Britain and Ireland, Harrogate, UK, May 2014; published in abstract form as *Br J Surg* 2013; **100**(Suppl 8): 44, *Br J Surg* 2014; **101**(Suppl 4): 18 and *Br J Surg* 2015; **102**(Suppl 1): 34

Paper accepted 23 June 2015
Published online 7 September 2015 in Wiley Online Library (www.bjs.co.uk). **DOI:** 10.1002/bjs.9905

Introduction

There is consensus internationally regarding initial staging algorithms for oesophageal cancer. For patients in whom potentially curative treatment may be both feasible and appropriate, the Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland (in conjunction with the British Society of Gastroenterology and British Association of Surgical Oncology), the USA Society of Thoracic Surgeons and National Comprehensive Cancer Network, and the European Society of Medical Oncology all recommend CT and endoscopic ultrasonography (EUS) for all patients, followed by PET–CT for T1b–T4 disease[1–4]. In the USA and Europe, laparoscopy is reserved mainly for T3–T4 disease involving the gastro-oesophageal junction (GOJ), although in the UK

it is also recommended for distal oesophageal tumours. These recommendations reflect the complementary roles of each modality: EUS is most accurate for assignment of T and N category[5], PET–CT for distant metastases[6,7] and laparoscopy for peritoneal disease[8].

Use of all modalities, however, consumes resources, delays treatment, and may confer small but clinically significant risks, while not necessarily altering management decisions. The primary utility of staging modalities is to divide patients into optimal management groups, identifying patients with early (T1a N0 intramucosal cancer for endoscopic resection (ER); T1b N0 for resection of submucosal cancer without neoadjuvant therapy), locally advanced (T1b N1 to T4a; typically resection with neoadjuvant therapy), and unresectable T4b and metastatic disease. Use of all modalities provides optimal staging precision and useful prognostic information (for example, distinguishing between T2 N1 and T3 N3 disease), but typically without altering management. It is not known whether patient subgroups exist for which these primary test utilities differ. Similarly, decision analytic metrics are not available to guide staging (whether this utility justifies test risk).

The primary aim of the present study was to calculate the net benefits and risks of EUS, PET–CT and laparoscopy, their primary utilities (probability of altering management) and probability thresholds (P_t; at which test benefit equals risk), using decision theory in a development data set. A secondary aim was to determine whether clinical, radiological and histopathological factors could be identified that were related to these endpoints, in order to generate predictive models to identify patient subgroups for selective staging. The final aim was then to refine existing staging algorithms on the basis of optimal pragmatism, maximal efficiency and minimal patient risk, evaluated using a validation data set.

Methods

Consecutive patients with oesophageal/GOJ cancer staged beyond CT in a UK centre were identified from four parallel databases. These patients represented those either diagnosed at the centre, or referred to the central oesophagogastric multidisciplinary team from three regional referring National Health Service (NHS) hospitals. Patients staged between May 2006 and July 2013 comprised a development set, and those staged from July 2013 to July 2014 an independent validation set. All investigations were reported and reviewed by a specialist multidisciplinary team, in accordance with the contemporary American Joint Committee on Cancer TNM staging manuals (6th[9] or 7th[10] edition)[8]. Patients without unequivocal metastases on CT were routinely staged sequentially using [^{18}F]fluorodeoxyglucose (FDG) PET–CT, EUS and laparoscopy, with oesophagogastroduodenoscopy (OGD) for GOJ tumours and distal oesophageal tumours extending below the diaphragm. Investigations were performed out of sequence in a small number of patients to maximize expediency. Neoadjuvant chemotherapy was considered for disease beyond T1 N0; ER was used from 2008 for possible T1a tumours. The study was registered with the Oxford University Hospitals NHS Trust clinical governance department as a service provision audit (number 2516), and as such was exempt from ethical approval.

CT was performed locally (367) or regionally (586) at three referring hospital trusts using a standard protocol[8] (11 scans were performed elsewhere). [^{18}F]FDG PET–CT was carried out using a General Electric Discovery STE 16-slice instrument (General Electric Healthcare, Milwaukee, Wisconsin, USA) (scanner 1; 60 min after 400 MBq [^{18}F]FDG) before November 2009, and thereafter a Discovery 690 64-slice system (scanner 2; 90 min after 4 MBq/kg [^{18}F]FDG), without intravenous contrast using standard iterative reconstruction, and reported independently by two dedicated PET–CT radiologists. EUS and ER were performed locally using 5–10-MHz radial or linear echoscopes (Hitachi, Wellingborough, UK), by one of three consultants and reported using the contemporary TNM edition and British Society of Gastroenterology guidelines[11]. Miniprobe EUS was performed using 20-Hz high-frequency catheter probes (Olympus UM-2R; Olympus, Southend-on-Sea, UK). ER was achieved using multiband ligation mucosectomy (Duette^{TM}; Cook Medical, Limerick, Ireland). Nodal stage was determined by ultrasonography, with fine-needle aspiration (FNA) if the result was considered likely to alter management. Potential bias was assessed for referring centre and PET scanner, but not EUS operator, as possible T1 N0 tumours were referred selectively for EUS with or without ER. At laparoscopy, visual examination of the peritoneal cavity was carried out without opening the lesser sac. Lavage cytology was not undertaken routinely.

Data collected were: patient age at diagnosis, sex, endoscopic findings, radiological findings (CT, EUS, PET–CT TNM staging; PET–CT maximum standardized uptake value (SUVmax), length, presence of avid nodes; additional PET–CT findings; subsequent investigations), pathological (pretreatment cell type; grade[10]; pathological stage, grade) and multidisciplinary team management decisions. Two TNM editions were used within this period; N status and overall stage were classified using the sixth edition as insufficient data were available for conversion to the seventh edition. Management was uniform for both (for

example, patients with M1a nodes based on the 6th edition were considered for radical treatment).

The primary utility of EUS was considered to be that altering management by identifying or refuting T1 N0 or T4b disease, or unsuspected metastases (distant organ, or nodal outside a standard resection field). For PET–CT, this was by identifying and confirming metastases, or unsuspected lesions altering management following appropriate specific investigation. The latter included treatment of a synchronous cancer or detection of high-risk lesions such as colonic polyps requiring surveillance (for example 3–4 adenomatous polyps, or those larger than 1 cm[12]). For laparoscopy, this was the identification of unresectable disease, either metastatic or T4b, with histopathological confirmation.

Statistical analysis

Analysis was performed using R version 3.0.2[13]. $P < 0{\cdot}050$ was corrected for multiple comparisons by the Bonferroni method[14]. Multivariable logistic regression included all variables following exclusion of perfect separators. Continuous variables were assessed (kernel density plots) and transformed (age^2; logSUVmax).

Decision analytic measures and cost analysis

P_t values were calculated to inform decision-making and models[15] (*Appendix S1*, supporting information[16–31]). Decision curve analysis (DCA) was performed to calculate the effect of varying P_t (0·496–2·719 per cent) reflecting EUS perforation risk (0·02–1·00 per cent)[32].

Model and algorithm development, validation and performance

Modelling was undertaken to corroborate proposed algorithms for selective EUS, PET–CT and laparoscopy, and to determine whether unidentified interactions could refine them. Three modelling techniques were used: logistic regression models (LRMs; backwards stepwise binary logistic), decision tree analysis (DTA; recursive partitioning) and artificial neural networks (ANNs; feed-forward back-propagation multilayer perceptron; *Appendix S1*, supporting information). Internal validation was carried out using bootstrapping/random forests (*Appendix S1*, supporting information); temporal validation was undertaken using the temporal validation data set.

Performance was quantified using calibration (observed : expected ratio), accuracy (Brier and κ scores), discrimination (area under the receiver operating characteristic (ROC) curve), and utility (sensitivity, specificity, positive (PPV) and negative (NPV) predictive values, and Peirce's net benefit[33]). Sensitivities and specificities were compared using McNemar's t test (DTComPair version 1.0.3[31]).

Results

A total of 953 consecutive patients were staged following CT by PET–CT (918), EUS (798) and laparoscopy (458). Of these, 829 comprised the development data set (800, 698 and 397 respectively) and 124 the validation set (118, 100 and 61) (*Table 1*).

PET–CT

Some 800 PET–CT examinations were performed in the development data set. Previous CT had suggested metastatic disease in 100 of these patients. This was confirmed in 57 (57·0 per cent) and refuted in 43 (43·0 per cent). In 28 (49 per cent) of the 57 patients with confirmed metastases, additional unsuspected metastases were demonstrated. Metastases were identified in 104 (14·9 per cent) of the 700 patients without a previous suggestion of metastatic disease, In total, therefore, metastatic disease was demonstrated in 161 patients (144 unequivocal, 17 requiring confirmatory biopsy). Five further lesions were biopsied, which demonstrated benign disease (4) and chronic lymphocytic leukaemia (1).

Some 119 non-metastatic lesions were identified in 103 patients. Seventy-eight patients were investigated, confirming significant pathology in 17 (2·1 per cent): nine synchronous cancers involving the thyroid (2), colon (2), lung (2), prostate (2) and bladder (1), and eight patients with high-risk colonic polyps. In six patients (0·8 per cent), PET–CT also staged known synchronous cancers. Consequently, PET–CT altered management in 23·0 per cent: confirming metastases (7·1 per cent), identifying unsuspected metastases (13·0 per cent) and additional pathology (2·1 per cent), and staging synchronous cancers (0·8 per cent).

Higher SUVmax was associated with squamous cell carcinoma (SCC), advanced stage, female sex and younger age; longer length with advanced stage; and FDG-avid nodes with SCC, advanced stage and age. The later PET–CT scanner was associated with a marginally higher SUVmax and avid nodes, but all reported associations were independent of this (*Tables S1–S3*, supporting information).

Predicting unsuspected metastases

Analysis was restricted to the 700 patients with CT M0 examinations (*Table S4*, supporting information), to

Table 1 Patient and tumour characteristics

	Development set ($n = 829$)	Validation set ($n = 124$)
Age (years)*§	66 (13, 29–88)	68 (13, 40–80)
Sex ratio (F : M)	204 : 625	35 : 89
Cell type		
Adenocarcinoma	639 (77·1)	104 (83·9)
Squamous cell carcinoma	165 (19·9)	18 (14·5)
Adenosquamous carcinoma	7 (0·8)	2 (1·6)
Neuroendocrine carcinoma	3 (0·4)	0 (0)
Small cell carcinoma	3 (0·4)	0 (0)
Undifferentiated	12 (1·4)	0 (0)
Grade		
Well	71 (8·6)	5 (4·0)
Moderate	347 (41·9)	58 (46·8)
Poor	389 (46·9)	60 (48·4)
Undifferentiated	22 (2·7)	1 (0·8)
Tumour site		
Proximal	23 (2·8)	1 (0·8)
Mid	67 (8·1)	12 (9·7)
Distal	237 (28·6)	42 (33·9)
Siewert 1	197 (23·8)	38 (30·6)
Siewert 2	195 (23·5)	16 (12·9)
Siewert 3	105 (12·7)	14 (11·3)
Multifocal	5 (0·6)	1 (0·8)
Pretreatment stage (TNM 6th edition)		
I	52 (6·3)	11 (8·9)
IIA	157 (18·9)	21 (16·9)
IIB	57 (6·9)	15 (12·1)
III	373 (45·0)	55 (44·4)
IV	190 (22·9)	22 (17·7)
Impassable at OGD		
No	699 (84·3)	107 (86·3)
Yes	130 (15·7)	17 (13·7)
Management decisions		
Palliative (performance status)	143 (17·2)	12 (9·7)
Palliative (incurable)	185 (22·3)	22 (17·7)
Endoscopic therapy	26 (3·2)	11 (8·9)
Surgery	58 (7·0)	5 (4·0)
Neoadjuvant therapy + surgery	378 (45·6)	67 (54·0)
DCRT	37 (4·5)	7 (5·6)
Died during staging	2 (0·2)	0 (0)
Endoscopic ultrasonography		
Not requested	127 (15·3)†	24 (19·4)‡
Performed (passable)	543 (65·5)	75 (60·5)
Performed (impassable + miniprobe)	49 (5·9)	1 (0·8)
Performed (impassable, no miniprobe)	106 (12·8)	24 (19·4)
Abandoned	4 (0·5)	0 (0)

Values in parentheses are percentages unless indicated otherwise; *values are median (i.q.r., range). Endoscopic ultrasonography not requested owing to: †metastases on PET–CT (79) or laparoscopy (5), patient choice (22), prohibitive anatomy (10), decision to proceed to radical chemoradiotherapy irrespective of findings (7), death during staging (2) and unclear reasons (2); ‡metastases on PET–CT (10) or laparoscopy (1), patient choice/performance status (9), prohibitive anatomy (1) and decision to proceed to definitive chemoradiotherapy (DCRT) (3). OGD, oesophagogastroduodenoscopy. §$P < 0·001$ (Shapiro–Wilk test).

identify variables associated with the demonstration of metastases on PET–CT. Although advancing EUS T and CT N category independently predicted metastases, no factors could be used to identify patients with a probability below the *P*t (0·083 per cent), that is patients in whom the risk of demonstrating metastases was sufficiently low not to justify the risk of PET–CT. Although there was zero incidence in EUS T1 disease, the 95 per cent c.i. was broad (0–6·12 per cent), suggesting that, contrary to common clinical practice, PET–CT may have utility in tumours staged by EUS as T1.

Endoscopic ultrasonography

T1 N0

In the development study, Tx/T1 N0 disease was reported in 51 (7·3 per cent) of 698 patients. In four tumours impassable to the standard echoscope, possible T1 N0 disease was demonstrated in the visualized tumour. However, in none was T1 N0 status confirmed by miniprobe. T1 N0 disease was associated with earlier CT T category, CT N0 status and a passable tumour at OGD (*Table 2*).

Of 128 patients with possible T1 disease on CT, 49 (38·3 (95 per cent c.i. 30·0 to 46·9) per cent) were identified by EUS, and two of 501 with CT T2–T4a disease (0·4 (0 to 1·5) per cent). There was zero incidence in the 69 patients with possible T4b disease on CT (95 per cent c.i. 0 to 4·4 per cent) and 81 with impassable tumours at OGD (95 per cent c.i. 0 to 3·7 per cent). Multivariable regression was therefore not possible for impassable tumours. Earlier CT T category independently predicted EUS T1 N0 disease (*Table 2*).

Some 675 of 698 patients subsequently underwent PET–CT (including 38 patients staged by EUS as T1 N0). There was zero incidence of EUS T1 N0 disease among the 259 patients with FDG-avid nodes (95 per cent c.i. 0 to 10·9 per cent). In addition to this, multivariable regression demonstrated shorter avid length and earlier CT T category to predict T1 N0 disease.

T4b

Among 69 patients with T4b disease suggested on CT, this was confirmed in 26 (38 per cent) patients (4 miniprobe EUS) and refuted in 31 (3 miniprobe EUS). A further 12 patients with impassable tumours were managed as T3. No additional patients were identified. On multivariable analysis (unadjusted for CT T category as a perfect separator) cell type (SCC and adenosquamous carcinoma), site (non-GOJ tumour), impassability and longer PET length independently predicted T4b disease (*Table 3*).

Table 2 Factors associated with T1 N0 disease on endoscopic ultrasonography: overall and multivariable logistic regression

	EUS T1 N0 ($n = 51$)	EUS ≥ T1 N1 ($n = 647$)	P (corrected $\alpha = 0{\cdot}0042$)§	Multivariable odds ratio†	P
Age (years)*‡	69 (59–72, 32–84)	66 (64–75, 29–88)	0·019¶	1·00 (1·00, 1·00)	0·649
Sex			1·000#		
F	12 (24)	158 (24·4)		Reference	Reference
M	39 (76)	489 (75·6)		0·70 (0·18, 2·80)	0·617
Cell type			0·209		
Adenocarcinoma	46 (90)	494 (76·4)		Reference	Reference
Squamous cell carcinoma	5 (10)	133 (20·6)		0·21 (0·02, 2·20)	0·192
Adenosquamous carcinoma	0 (0)	7 (1·1)		n.a.	n.a.
Neuroendocrine carcinoma	0 (0)	3 (0·5)		n.a.	n.a.
Small cell carcinoma	0 (0)	2 (0·3)		n.a.	n.a.
Undifferentiated	0 (0)	8 (1·2)		n.a.	n.a.
Grade			0·525		
Well	6 (12)	59 (9·1)		Reference	Reference
Moderate	25 (49)	269 (41·6)		2·54 (0·32, 20·10)	0·367
Poor	19 (37)	302 (46·7)		3·03 (0·37, 24·76)	0·302
Undifferentiated	1 (2)	17 (2·6)		0·00 (0, ∞)	0·996
Tumour site			0·050		
Proximal	0 (0)	14 (2·2)		n.a.	n.a.
Mid	4 (8)	56 (8·7)		Reference	Reference
Distal	21 (41)	173 (26·7)		0·71 (0·03, 14·71)	0·822
Siewert 1	17 (33)	160 (24·7)		1·67 (0·07, 39·77)	0·750
Siewert 2	8 (16)	160 (24·7)		0·42 (0·02, 9·88)	0·591
Siewert 3	1 (2)	80 (12·4)		2·16 (0·50, 9·25)	0·534
Multifocal	0 (0)	4 (0·6)		n.a.	n.a.
Impassable (OGD)			0·002		
No	51 (100)	566 (87·5)		n.a.	n.a.
Yes	0 (0)	81 (12·5)			
CT T category			< 0·001		
Possible T1	49 (96)	79 (12·2)		Reference	Reference
T2–T4a	2 (4)	499 (77·1)		0·03 (0·00, 0·16)	< 0·001
Possible T4b	0 (0)	69 (10·7)		n.a.	n.a.
CT N category			< 0·001		
N0	49 (96)	274 (42·3)		Reference	Reference
N1	2 (4)	373 (57·7)		0·75 (0·08, 7·39)	0·803
CT centre			0·954		
OUH	26 (51)	226 (34·9)		2·61 (0·62, 10·95)	0·188
Centre 1	8 (16)	157 (24·3)		Reference	Reference
Centre 2	13 (25)	150 (23·2)		2·16 (0·50, 9·25)	0·301
Centre 3	1 (2)	92 (14·2)		0·44 (0·03, 6·24)	0·542
Other	3 (6)	22 (3·4)		29·4 (0·58, 1503·75)	0·100
SUVmax*	4·15 (2·50–5·13, 2·50–11·80)	10·50 (7·08–15·9, 2·50–52·60)	< 0·001¶	0·06 (0·00, 3·37)	0·168
Not performed	13 (25)	10 (1·5)			
PET length (cm)*	1·00 (0–1·78, 0–6·90)	5·40 (3·70–7·00, 0–20·00)	< 0·001¶	0·53 (0·29, 0·95)	0·034
Not performed	13 (25)	10 (1·5)			
PET avid nodes			< 0·001	n.a.	n.a.
No	38 (75)	376 (58·1)			
Yes	0 (0)	259 (40·0)			
Not performed	13 (25)	12 (1·9)			
PET scanner			n.a.		
1	14 (27)	304 (47·0)		Reference	Reference
2	24 (47)	331 (51·2)		2·11 (0·67, 6·66)	0·203
Not performed	13 (25)	12 (1·9)			

Values in parentheses are percentages unless indicated otherwise; values are *median (i.q.r., range) and †95 per cent c.i. The analysis was done following exclusion of perfect separators. EUS, endoscopic ultrasonography; n.a., not applicable; OGD, oesophagogastroduodenoscopy; OUH, Oxford University Hospitals; SUVmax, maximum standardized uptake value. ‡$P < 0{\cdot}001$ (Shapiro–Wilk test). §Pearson's χ^2 test, except ¶Mann–Whitney U test and #Fisher's exact test.

Table 3 Factors associated with T4b disease on endoscopic ultrasonography: overall and binary logistic regression

	EUS T4b ($n = 26$)	EUS < T4b ($n = 672$)	P (corrected $\alpha = 0{\cdot}0042$)§	Multivariable odds ratio†	P
Age (years)*‡	67·5 (58·3–69·8, 48·0–80·0)	66 (59·8–72·3, 44·0–83·0)	0·475¶	1·00 (1·00, 1·00)	0·790
Sex			0·059#		
F	2 (8)	168 (25·0)		Reference	Reference
M	24 (92)	504 (75·0)		0·49 (1·72, 1·40)	0·182
Cell type			< 0·001		
Adenocarcinoma	7 (27)	533 (79·3)		Reference	Reference
Squamous cell carcinoma	17 (65)	121 (18·0)		4·98 (1·27, 19·49)	0·021
Adenosquamous carcinoma	1 (4)	6 (0·9)		54·1 (2·68, 1091·91)	0·009
Neuroendocrine carcinoma	0 (0)	3 (0·4)		n.a.	n.a.
Small cell carcinoma	1 (4)	1 (0·2)		n.a.	n.a.
Undifferentiated	0 (0)	8 (1·2)		n.a.	n.a.
Grade			< 0.001		
Well	4 (15)	61 (9·1)		Reference	Reference
Moderate	12 (46)	282 (42·0)		0·25 (0·05, 1·18)	0·081
Poor	8 (31)	313 (46·6)		0·20 (0·04, 1·10)	0·065
Undifferentiated	2 (8)	16 (2·4)		0·00 (0, ∞)	0·991
Tumour site			< 0·001		
Proximal	2 (8)	12 (1·8)		Reference	Reference
Mid	12 (46)	48 (7·1)		1·24 (0·20, 7·89)	0·865
Distal	7 (27)	187 (27·8)		0·31 (0·05, 2·07)	0·228
Siewert 1	3 (12)	174 (25·9)		0·05 (0·00, 0·68)	0·025
Siewert 2	0 (0)	168 (25·0)		n.a.	n.a.
Siewert 3	1 (4)	80 (11·9)		0·13 (0·01, 2·30)	0·164
Multifocal	1 (4)	3 (0·5)		0·52 (0·02, 14·68)	0·699
Impassable (OGD)			0·002		
No	17 (65)	600 (89·3)		Reference	Reference
Yes	9 (35)	72 (10·7)		4·15 (1·31, 13·12)	0·015
CT T category			< 0·001		
Possible T1	0 (0)	128 (19·0)		n.a.	n.a.
T2–T4a	0 (0)	501 (74·6)		n.a.	n.a.
Possible T4b	26 (100)	43 (6·4)		n.a.	n.a.
CT N category			1·000#		
N0	12 (46)	311 (46·3)		Reference	Reference
N1	14 (54)	361 (53·7)		0·90 (0·30, 2·70)	0·846
CT centre			0·664		
OUH	9 (35)	243 (36·2)		0·44 (0·12, 1·57)	0·205
Centre 1	9 (35)	156 (23·2)		Reference	Reference
Centre 2	4 (15)	159 (23·7)		0·22 (0·04, 1·06)	0·059
Centre 3	3 (12)	90 (13·4)		0·44 (0·08, 2·51)	0·358
Other	1 (4)	24 (3·6)		1·42 (0·12, 17·60)	0·785
SUVmax*	13·80 (9·70–16·80, 2·90–32·50)	10·30 (7·00–15·90, 2·50–52·60)	0·062¶	0·36 (0·12, 1·11)	0·074
PET length*	7·00 (5·93–8·28, 2·00–14·40)	5·00 (3·50–7·00, 0–20·00)	< 0·001¶	1·63 (1·29, 2·06)	< 0·001
PET avid nodes					
No	12 (46)	402 (59·8)	0·337#	Reference	Reference
Yes	12 (46)	247 (36·8)		1·03 (0·32, 3·34)	0·965
Not performed	2 (8)	23 (3·4)			
PET scanner			n.a.		
1	17 (65)	301 (46·5)		Reference	Reference
2	7 (27)	348 (53·8)		0·36 (0·12, 1·10)	0·074
Not performed	2 (8)	23 (3·6)			

Values in parentheses are percentages unless indicated otherwise; values are *median (i.q.r., range) and †95 per cent c.i. The analysis was done following exclusion of perfect separators. EUS, endoscopic ultrasonography; n.a., not applicable; OGD, oesophagogastroduodenoscopy; OUH, Oxford University Hospitals; SUVmax, maximum standardized uptake value. ‡$P < 0.001$ (Shapiro–Wilk test). §Pearson's χ^2 test, except ¶Mann–Whitney U test and #Fisher's exact test.

Table 4 Apparent, internally validated and temporally validated performance metrics of existing and novel algorithms, plus decision tree analyses, logistic regression models and artificial neural networks in predicting T1 N0, T4b and pT1 N0 disease by endoscopic ultrasonography

Algorithm/model	Sensitivity	Specificity	PPV	NPV	Brier	κ	O : E	AUC	Net benefit	Total cost (€)
EUS T1 N0 ($n = 51$ development; $n = 11$ temporal validation)										
Default	1·000	0·000	0·738	0·000	n.a.	0·000	0·731	0·500	0·050	534 803
Novel (apparent)	0·961	0·878	0·383	0·997	0·046	0·495	0·398	0·919	0·071	118 671
Novel (internal)	0·941	0·881	0·384	0·995	0·137	0·493	0·349	0·928	0·067	128 971
Novel (independent)	1·000	0·855	0·478	1·000	0·048	0·581	0·478	0·928	0·107	n.a.
Models before PET–CT										
DTA 1 (apparent)	0·961	0·878	0·383	0·997	0·046	0·495	0·398	0·919	0·071	118 671
DTA 1 (internal)	0·941	0·881	0·384	0·995	0·137	0·493	0·349	0·928	0·070	128 971
DTA 1 (independent)	1·000	0·855	0·478	1·000	0·048	0·581	0·478	0·928	0·107	n.a.
LRM1 (apparent)	0·980	0·535	0·157	0·997	0·055	0·152	0·160	0·758	0·068	324 243
LRM 1 (internal)	0·980	0·535	0·157	0·997	0·055	0·152	0·160	0·758	0·068	324 243
LRM 1 (independent)	1·000	0·723	0·324	1·000	0·052	0·379	0·324	0·861	0·107	n.a.
ANN 1 (apparent)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
ANN 1 (internal)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
ANN 1 (independent)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Models after PET–CT										
DTA 2 (apparent)	1·000	0·840	0·331	1·000	0·031	0·436	0·331	0·920	0·072	159 368
DTA 2 (internal)	0·944	0·866	0·225	0·996	0·038	0·401	0·298	0·907	0·051	179 967
LRM 2 (apparent)	0·947	0·734	0·310	0·991	0·049	0·359	0·301	0·953	0·100* (0·069)	179 967
LRM 2 (internal)	0·947	0·734	0·310	0·991	0·052	0·036	0·301	0·946	0·100* (0·069)	179 967
ANN 2 (apparent)	0·974	0·835	0·261	0·998	0·022	0·354	0·261	0·904	0·069	169 667
ANN 2 (internal)	0·974	0·835	0·261	0·998	0·018	0·354	0·261	0·904	0·068	169 667
EUS T4b ($n = 26$ development; $n = 0$ temporal validation)										
Default	1·000	0·000	0·038	0·000	n.a.	n.a.	n.a.	0·500	0·009	534 803
Novel (apparent)	1·000	0·936	0·377	1·000	0·038	0·522	0·377	0·968	0·035	52 867
Novel (internal)	0·962	0·926	0·333	0·998	0·029	0·466	0·333	0·944	0·034	91 995
DTA (apparent)	1·000	0·936	0·377	1·000	0·038	0·522	0·377	0·968	0·035	52 867
DTA (internal)	0·962	0·926	0·333	0·998	0·029	0·466	0·333	0·944	0·034	91 995
ANN (apparent)	1·000	0·700	0·114	1·000	0·019	0·147	0·114	0·849	0·028	52 867
ANN (internal)	1·000	0·702	0·115	1·000	0·085	0·147	0·115	0·851	0·028	52 867
pT1 N0										
Default	1·000	0·000	0·564	0·000	n.a.	n.a.	n.a.	0·500	0·466	n.a.
DTA (apparent)	1·000	0·348	0·580	1·000	0·200	0·333	0·580	0·790	0·470	n.a.
DTA (internal	0·667	0·708	0·741	0·630	0·421	0·370	0·889	0·685	0·308	n.a.
LRM (apparent)	1·000	0·111	0·590	1·000	0·102	0·139	0·583	0·556	0·440	n.a.
LRM (internal)	1·000	0·111	0·590	1·000	0·102	0·139	0·583	0·556	0·440	n.a.
ANN (apparent)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
ANN (internal)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

*LRM net benefit derived on smaller subset; adjusted benefit for whole cohort in parentheses. PPV, positive predictive value; NPV, negative predictive value; 0 : E, observed : expected; AUC, area under curve; EUS, endoscopic ultrasonography; n.a., not applicable; internal, internal validation; independent, independent validation; DTA, decision tree analysis; LRM, logistic regression model; ANN, artificial neural network.

Metastases

Possible M1 nodal metastases were identified by EUS in five patients; they were excluded by FNA cytology in three and confirmed in two (both evident on PET–CT).

Utility

EUS altered management decisions in 77 patients (11·0 per cent). Although PET–CT was typically performed before EUS, in a limited number it was performed on the same or subsequent days. Metastases were demonstrated in 96 patients (14·2 per cent) by further staging with PET–CT with or without laparoscopy. In patients downstaged by EUS from possible T4b disease, metastases on PET–CT were subsequently demonstrated in ten of 43 patients.

Stratification by CT T category demonstrated distinct utilities: in patients with Tx/possible T1 (early) disease (128, 18·3 per cent), EUS confirmed T1 N0 in 49 (38·3 per cent). In patients with possible T4b disease (69, 9·9 per cent) EUS was confirmatory in 26 (38 per cent). However, in 501 patients (71·8 per cent) without possible T1 or T4b disease on CT, EUS altered management in just two (0·4 per cent). In the 81 patients with impassable tumours, EUS altered management in three (4 per cent), confirming T4b with miniprobe EUS.

Comparison with pathological staging

Among 367 patients who underwent resection, 81 (22·1 per cent) received no neoadjuvant therapy, of whom 19 underwent ER, 7 ER plus surgery and 55 surgery. Of the 51 EUS T1 N0 tumours, 46 were resected (17, 7 and 22 respectively). Of the 17 patients who had ER alone, pT1 disease was confirmed in 11. Among the 29 treated with surgery with or without ER, pT1 N0 disease was confirmed in 24. Five tumours were upstaged by T category, two also to N1 disease. Five other tumours staged by EUS as greater than T1 N0 were downstaged to pT1 N0, from EUS T3 (3) and EUS T1 N1 (2).

Excluding the 17 patients who, after EUS, underwent ER without surgical resection (in whom pN status could not be assessed), EUS was 83 per cent sensitive and 84 per cent specific for pT1 N0 (PPV 83 per cent; NPV 84 per cent). Factors associated with pT1 N0 disease are presented in *Table S5* (supporting information).

Endoscopic ultrasonography and decision theory

There was one instance of perforation (0·1 per cent). The P_t for EUS T1 N0 disease was 0·95 per cent (the probability of identifying T1 N0 disease at which the benefit of EUS equals its risk). As the probability in patients with impassable tumours (0 per cent) or T2–T4a disease on CT (0·4 per cent) was lower, the risk of EUS to these patients outweighed its potential benefit of altering management. The P_t for EUS T4b disease was 2·02 per cent (based on T4 disease overall).

Staging laparoscopy

Some 397 patients underwent laparoscopy, and metastases were demonstrated in 28 (7·1 per cent). In 341 patients undergoing laparoscopy without feeding jejunostomy there was one major complication (pneumonia, 0·3 per cent) and four minor complications (1·2 per cent, urinary retention). Metastases were demonstrated in two (4 per cent) of 54 distal oesophageal tumours not involving the GOJ endoscopically. An impassable tumour, undifferentiated grade, possible T4b disease on CT, and lower SUVmax predicted unsuspected peritoneal metastases (*Table S6*, supporting information). No factor could identify patients below the P_t (0·38 per cent).

Refinement of existing algorithm

As a result of the findings that the incidence of T1 N0 disease on EUS among patients staged as T2–T4a by CT was minimal, and insufficient to justify the EUS test risk, it is proposed that EUS should be reserved only for patients with possible T1 or T4b disease on CT. This would have been 96·1 per cent sensitive and 87·8 per cent specific for T1 N0, and 100 per cent sensitive and 93·6 per cent specific for T4b disease on EUS, with very high NPV (99·7 and 100 per cent respectively) (*Table 4*).

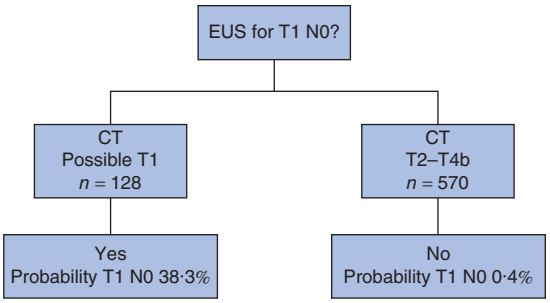

Fig. 1 CT-guided algorithm (and decision tree analysis 1) for performing endoscopic ultrasonography (EUS) for T1 N0 disease before PET–CT. Numbers relate to development set

Compared with existing practice there was no significant change in sensitivity, but specificity improved ($P < 0·001$) and net benefits increased. This would be accompanied by a 71·8 per cent reduction in examinations, procedural risk and expenditure. DCA demonstrated minimal change in net algorithm benefits with extreme variations in perforation risk (*Fig. S1*, supporting information).

The zero incidence of T1 N0 disease in impassable tumours also suggests that EUS could be omitted in these patients. However, this would have applied to only two of 128 patients with possible T1 disease on CT. Similarly, the zero incidence in patients with FDG-avid nodes applied to only 17 patients with FDG-avid nodes staged by CT as possible T1 disease, for whom confirmatory assessment by EUS with or without FNA may be warranted (*Fig. 1*)

This CT-guided algorithm was validated internally (using bootstrapping) and independently. Internal validation demonstrated minimal overfitting and sampling bias; the adjusted T1 N0 sensitivity was 94·1 per cent and the specificity 88·1 per cent; for T4b, respective values were 96·2 and 92·6 per cent (*Table 4*). Independent validation demonstrated 100·0 per cent sensitivity and 85·5 per cent specificity for T1 N0 (11 EUS T1 N0).

Validation of new endoscopic ultrasonography algorithm

Some 91 patients in the validation set underwent PET–CT and EUS. No patient was staged by EUS as having T1 N0 disease among the 60 with avid nodes. Twelve had possible T4b disease on CT; seven underwent EUS refuting T4b and EUS was omitted in five. EUS was performed in

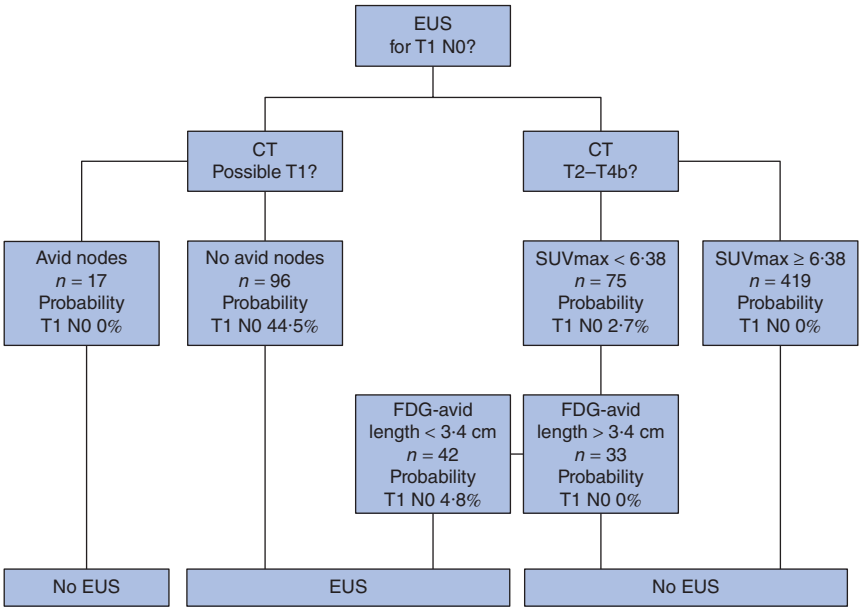

Fig. 2 Decision tree analysis 2 for performing endoscopic ultrasonography (EUS) for T1 N0 disease after PET–CT. SUVmax, maximum standardized uptake value. Numbers relate to development set

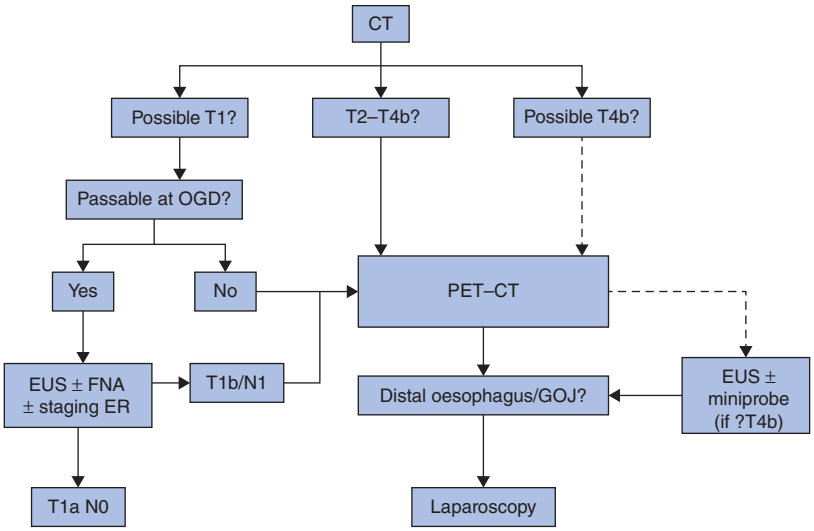

Fig. 3 Pragmatic algorithm for staging oesophageal cancer. OGD, oesophagogastroduodenectomy; EUS, endoscopic ultrasonography; FNA, fine-needle aspiration; ER, endoscopic resection; GOJ, gastro-oesophageal junction

eight of 17 patients with impassable tumours; two were downstaged from possible T4b disease.

Modelling

The optimal model for identifying T1 N0 disease by EUS before PET–CT was a decision tree (DTA 1); this reserved EUS for those with possible T1 disease on CT, and was identical to the pragmatic CT-guided algorithm (*Fig. 1* and *Table 4*; *Table S7*, supporting information). After PET–CT, the optimal model was a modified decision tree (DTA 2); this reserved EUS for patients with possible T1 disease on CT without FDG-avid nodes, or CT T2–T4a disease with SUVmax below 6·38 and length less than 3·4 cm on PET–CT (*Fig. 2*). This was 100 per cent sensitive and 84·0 per cent specific with minimal overfitting on internal

validation (*Table 4*). This could not be validated independently, as no patient in the validation set with CT T2–T4a disease had EUS T1 N0 disease. DCA demonstrated variation in EUS perforation rate to have minimal effects on model net benefits (*Fig. S1*, supporting information)

Again, the optimal model for identifying T4b disease by EUS was a decision tree identical to the proposed algorithm; this reserved EUS for patients with possible T4b disease on CT (100 per cent sensitivity) (*Table 4*). ANNs could be generated, but no model had clinical utility. Similarly, no model had clinical utility in predicting pT1 N0 disease, and unsuspected metastases on PET–CT and at laparoscopy.

Suggested staging algorithm

Based on these findings, the following staging algorithm is proposed when considering patients for resection (*Fig. 3*). Following CT, EUS (with or without FNA or staging ER) should be reserved for patients with either: Tx/possible T1 disease on CT, passable at OGD; or possible T4b disease without metastases on PET–CT. For all other patients EUS can be omitted, thereby reducing risk, delay and expenditure.

Discussion

This study sought to quantify the staging utilities of EUS, PET–CT and laparoscopy for oesophageal/GOJ cancer, within the context of risk, benefit and probability thresholds. A number of factors predicted EUS T1 N0 and T4b disease, identifying patients above and below these thresholds. Although EUS provided additional precision and information regarding T and N categories over CT and PET–CT, this had no added utility in the majority of patients. For these patients, EUS risk exceeded potential benefit (as defined by its primary utility). This pragmatic algorithm was validated internally and independently.

The main findings regarding utilities of EUS, PET–CT and laparoscopy, and CT T category grouping and impassable tumours, are likely to be robust, owing to the size of the cohorts. Although the lower resolution of CT confers lower sensitivity and specificity for T and N category than EUS[34,35], it typically demonstrates early Tx/possible T1, possible T4b or locally advanced T2–T4a disease, and this appears sufficient to guide use of EUS. As the relative merits and results of investigations vary with local resources, imaging and software platforms, however, external evaluation of these algorithms is required to assess generalizability. These recommendations are based on pragmatic definitions of the primary utilities of EUS: the identification of patients for endoscopic resection and neoadjuvant therapy. The benefit of the latter in patients with T2 N0 disease has not been established unequivocally (as reflected by variations in study protocols). In selected patients, EUS may also have secondary utilities, such as providing additional prognostic or anatomical information (for example, distinguishing between T2 N0 and T3 N3 disease). Owing to the limitations of predicting outcome on the basis of EUS T and N categories before potentially downstaging neoadjuvant therapy, such information is likely to alter management infrequently. However, in such instances EUS may have a role to play in patients with T2–T4a disease on CT.

Although variables were associated with metastases on PET–CT and at laparoscopy, comparable groups could not be identified. No benefit was seen for PET–CT in the few tumours staged as T1 N0, but owing to the significant EUS false-positive rate (8·9 per cent) the possibility of metastases in this group exists, reinforcing the need for staging ER[4]. Indeed, PET variables could predict false-positive T1 N0 disease. Notably, there was benefit of laparoscopy for T2 and distal oesophageal tumours, which conflicts with American and European guidelines[1–3]. The study also confirmed that adenocarcinoma and SCC differ in their FDG avidity[8], and identified novel associations with FDG-avid nodes, SUVmax and length. The role of miniprobe EUS in impassable tumours remains unclear. There was benefit in patients with possible T4b disease on CT, but none for those without, although the 95 per cent c.i. for the latter (0 to 3·70 per cent) exceeded the P_t. Although the present study is retrospective, bias was minimized by collecting data from four parallel databases to ensure accuracy and that no data were missing. Having CT performed in multiple centres was not a significant confounder and was adjusted for. The advantages and disadvantages of DTA, LRM and ANN were mitigated by comparing all three.

On the basis of pragmatic primary staging utilities, the risk of EUS typically exceeded its benefit in patients with either T2–T4a disease on CT, impassable tumours at OGD or FDG-avid nodes on PET–CT. Laparoscopy was beneficial in patients with distal oesophageal tumours and T2 disease. There may be further roles for DTA and LRMs in guiding further selection of patients with oesophagogastric cancers for specific treatment pathways.

Acknowledgements

The authors thank R. Marshall, B. Sgromo and F. Gleeson for critical review of the manuscript; R. Styles for her work in coordinating the Oxford OesophagoGastric

multidisciplinary team and maintenance of its database; and D. Lunn for statistical advice.
Disclosure: The authors declare no conflict of interest.

References

1 Varghese TK Jr, Hofstetter WL, Rizk NP, Low DE, Darling GE, Watson TJ *et al*. The Society of Thoracic Surgeons guidelines on the diagnosis and staging of patients with esophageal cancer. *Ann Thorac Surg* 2013; **96**: 346–356.
2 National Comprehensive Cancer Network (NCCN). *NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines): Esophageal and Esophagogastric Junction Cancers. Version 1*; 2014. https://www.nccn.org/store/login/login.aspx?ReturnURL=http://www.nccn.org/professionals/physician_gls/pdf/esophageal.pdf [accessed 12 July 2017].
3 Stahl M, Budach W, Meyer HJ, Cervantes A; ESMO Guidelines Working Group. Esophageal cancer: Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2010; **21**(Suppl 5): v46–v49.
4 Allum WH, Blazeby JM, Griffin SM, Cunningham D, Jankowski JA, Wong R *et al*. Guidelines for the management of oesophageal and gastric cancer. *Gut* 2011; **60**: 1449–1472.
5 Puli SR, Reddy JB, Bechtold ML, Antillon D, Ibdah JA, Antillon MR. Staging accuracy of esophageal cancer by endoscopic ultrasound: a meta-analysis and systematic review. *World J Gastroenterol* 2008; **14**: 1479–1490.
6 Gillies RS, Middleton MR, Han C, Marshall RE, Maynard ND, Bradley KM *et al*. Role of positron emission tomography–computed tomography in predicting survival after neoadjuvant chemotherapy and surgery for oesophageal adenocarcinoma. *Br J Surg* 2012; **99**: 239–245.
7 Richardson JR, Khan OA. In patients with radiologically-staged resectable oesophago-gastric junctional tumours, is diagnostic laparoscopy useful as an additional staging procedure? *Int J Surg* 2012; **10**: 198–202.
8 Gillies RS, Middleton MR, Maynard ND, Bradley KM, Gleeson FV. Additional benefit of ^{18}F-fluorodeoxyglucose integrated positron emission tomography/computed tomography in the staging of oesophageal cancer. *Eur Radiol* 2011; **21**: 274–280.
9 Greene FL. The American Joint Committee on Cancer: updating the strategies in cancer staging. *Bull Am Coll Surg* 2002; **87**: 13–15.
10 Rice TW, Blackstone EH, Rusch VW. 7th edition of the AJCC Cancer Staging Manual: esophagus and esophagogastric junction. *Ann Surg Oncol* 2010; **17**: 1721–1724.
11 UK EUS Users Group, British Society of Gastroenterology. *Staging of Oesophago-gastric Carcinoma by Endoscopic Ultrasonography: Guidance and Minimum Standards*. British Society of Gastroenterology: London, 2004.
12 Cairns SR, Scholefield JH, Steele RJ, Dunlop MG, Thomas HJ, Evans GD *et al*. Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002). *Gut* 2010; **59**: 666–689.
13 R Foundation for Statistical Computing. *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, 2013.
14 Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995; **310**: 170.
15 Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980; **302**: 1109–1117.
16 Pech O, May A, Manner H, Behrans A, Pohl J, Weferling M *et al*. Long-term efficacy and safety of endoscopic resection for patients with mucosal adenocarcinoma of the esophagus. *Gastroenterology* 2014; **146**: 652–660.
17 Medical Research Council Oesophageal Cancer Working Group. Surgical resection with or without preoperative chemotherapy in oesophageal cancer: a randomised controlled trial. *Lancet* 2002; **359**: 1727–1733.
18 Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation; Board on Radiation Effects Research; Division on Earth and Life Studies; National Research Council. *Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII, Phase 2*. National Academies Press: Washington, DC, 2006.
19 Rouvelas I, Zeng W, Lindblad M, Viklund P, Ye W, Lagergren J. Survival after surgery for oesophageal cancer: a population-based study. *Lancet Oncol* 2005; **6**: 864–870.
20 Jamieson GG, Mathew G, Ludemann R, Wayman J, Myers JC, Devitt PG. Postoperative mortality following oesophagectomy and problems in reporting its rate. *Br J Surg* 2004; **91**: 2983–2990.
21 Findlay JM, Tilson RC, Harikrishnan A, Sqromo B, Marshall RE, Maynard ND *et al*. Attempted validation of the NUn score and inflammatory markers as predictors of esophageal anastomotic leak and major complications. *Dis Esophagus* 2014; [Epub ahead of print].
22 Nelson LM, Bloch DA, Longstreth WT Jr, Shi H. Recursive partitioning for the identification of disease risk subgroups: a case–control study of subarachnoid hemorrhage. *J Clin Epidemiol* 1998; **51**: 199–209.
23 Tu JV. Advantages and disadvantages of using artificial neural networks *versus* logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996; **49**: 1225–1231.
24 Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; **54**: 774–781.
25 Harrell FE Jr. *RMS: Regression Modeling Strategies. R package version 4.1-0*. R Foundation for Statistical Computing: Vienna, 2013. http://CRAN.R-project.org/package=rms [accessed 12 July 2015].
26 Therneau T, Atkinson B, Ripley, B. *rpart: Recursive Partitioning. R package version 4.1-4*. R Foundation for Statistical Computing: Vienna, 2013. http://CRAN.R-project.org/package=rpart [accessed 12 July 2015].
27 Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2002; **2**: 18–22.
28 Venables WN, Ripley BD. *Modern Applied Statistics with S* (4th edn). Springer: New York, 2002.

29 Kuhn M, Wing, J, Weston S, Williams A, Keefer C, Engelhardt A *et al. Caret: Classification and Regression Training. R package version 6.0-35*. R Foundation for Statistical Computing: Vienna, 2014. http://CRAN.R-project.org/package=caret [accessed 12 July 2015].
30 Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* 2003; **160**: 249–264.
31 Stock C, Hielscher, T. *DTComPair: Comparison of Binary Diagnostic Tests in a Paired Study Design. R package version 1.0.3*. R Foundation for Statistical Computing: Vienna, 2014. http://CRAN.R-project.org/package=DTComPair [accessed 12 July 2015].
32 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; **26**: 565–574.
33 Peirce CS. The numerical measure of the success of predictions. *Science* 1884; **4**: 453–454.
34 Lowe VJ, Booya F, Fletcher JG, Nathan M, Jensen E, Mullan B *et al*. Comparison of positron emission tomography, computed tomography, and endoscopic ultrasound in the initial staging of patients with esophageal cancer. *Mol Imaging Biol* 2005; **7**: 422–430.
35 van Vliet EP, Heijenbrok-Kal MH, Hunink MG, Kuipers EJ, Siersema PD. Staging investigations for oesophageal cancer: a meta-analysis. *Br J Cancer* 2008; **98**: 547–557.

Supporting information

Additional supporting information may be found in the online version of this article:

Appendix S1 Additional methods (Word document)

Table S1 Linear regression – factors associated with logSUVmax (Word document)

Table S2 Linear regression – factors associated with PET length (Word document)

Table S3 Binary logistic regression – factors associated with PET avid nodes (Word document)

Table S4 Factors associated with demonstration of unsuspected PET–CT metastases (binary logistic regression, excluding perfect separators) (Word document)

Table S5 Factors associated with pT1 N0 disease: binary logistic regression (Word document)

Table S6 Factors associated with metastases at staging laparoscopy: binary logistic regression (Word document)

Table S7 Statistical comparison of model performance (McNemar's *t* test for sensitivity and specificity) (Word document)

Fig. S1 Decision curve analysis for endoscopic ultrasonography T1 N0 models: initial models and including PET variables (Word document)