Assessment of the quality of surgery within randomised controlled trials for the treatment of gastro-oesophageal cancer: a systematic review

Sheraz R Markar, Tom Wiggins, Melody Ni, Ewout W Steyerberg, J Jan B Van Lanschot, Mitsuru Sasako, George B Hanna

Multicentre, randomised, controlled trials (RCTs) provide level 1 evidence for surgery in the treatment of gastro-oesophageal cancer. This systematic review investigated whether standardisation of surgical techniques in RCTs reduces the variation in lymph-node harvest, in-hospital mortality, and locoregional cancer recurrence. The range in the coefficients of variation for lymph-node harvest (0·07–0·61), proportion of patients with locoregional cancer recurrence (1·1–46·2%), and in-hospital mortality (0–10%) was wide. Credentialing of surgeons through assessment of operative reports and monitoring of their performance through data collection were important factors that reduced the variation in lymph-node harvest. Factors that reduced adjusted in-hospital mortality included credentialing surgeons through procedural volume and operative reports, and standardisation of surgical techniques. Future RCTs should include an assessment of surgical performance as an important aspect of study design to reduce variation in clinical outcomes.

Lancet Oncol 2015; 16: e23–31

Division of Surgery, Department of Surgery and Cancer, St Mary's Hospital, Imperial College, London, UK (S R Markar MRCS, T Wiggins MRCS, M Ni PhD, Prof G B Hanna PhD)**; Centre for Medical Decision Sciences, Department of Public Health** (Prof E W Steyerberg PhD)**, and Department of Surgery** (Prof J J B Van Lanschot PhD)**, Erasmus MC-University Medical Centre, Rotterdam, Netherlands; and Department of Surgery, Hyogo College of Medicine, Nishinomiya, Japan** (Prof M Sasako PhD)

Correspondence to:
Prof George Hanna, Division of Surgery, Department of Surgery and Cancer, Imperial College London, St Mary's Hospital, London W2 1NY, UK
g.hanna@imperial.ac.uk

Introduction

An estimated 482 300 new cases of oesophageal cancer and 406 800 oesophageal cancer deaths occur worldwide every year.[1] Gastric cancer affects nearly 1 million people globally every year and causes around 10% of all cancer deaths.[1] The mainstay of curative treatment for gastro-oesophageal cancer is surgery with resection of the tumour and relevant lymph nodes. However, during the past two decades, controversy has arisen regarding the management of gastro-oesophageal cancer, including the use of chemotherapy and radiotherapy, the extent of lymphadenectomy, and the effect of various surgical approaches. To investigate these areas, a large series of randomised controlled trials (RCTs) has been undertaken in several centres and international collaborations. These trials have provided level 1 quality evidence for clinical practice.

The frameworks for the assessment of any risk of potential bias within RCTs have been generated mainly for pharmacological trials[2–4] and are therefore not applicable for surgical trials with variability in technical performance. Currently, no definitive system is available to standardise and assess surgical performance within RCTs. Variation in surgical technique and performance within an RCT is likely to be equally distributed and might therefore not affect the overall conclusion of the investigation. This concept compromises the conclusion of such studies for two main reasons. First, if the intervention is the surgical technique, non-standardisation will lead to partial homogenisation of both arms of the study and will undermine the randomisation. The best example of such a situation is the Dutch D1 versus D2 gastrectomy trial,[5] in which 52% of operations in the D1 resection group had more widespread dissection than specified, and 84% of operations in the D2 gastrectomy group had less dissection than specified, which led to partial homogenisation of both groups, and reduced the probability of detecting any potential advantage to D2 dissection. Second, if the intervention is an oncological method, non-standard surgical techniques will change the effect of the intervention, leading to variable outcomes in different groups. For instance, in a subgroup analysis of an RCT by MacDonald and colleagues[6] comparing postoperative chemoradiotherapy versus surgery alone, adjuvant chemoradiotherapy was needed after D0–1 resection, but it had no added value after D2 gastrectomy.

Reported RCTs investigating oesophageal and gastric cancer surgery have used variable methods or no methods at all to ensure the quality of surgical performance. No work has been done to examine the effect of variability in technical performance on the outcomes of RCTs. Therefore, in this Review, we aimed to identify important factors in the design of surgical RCTs that could reduce any bias in trial results. These factors will provide the basis for a proposed method to assure the quality of surgical RCTs. This Review investigated whether standardisation of surgical techniques in RCTs would first, produce more reliable results than non-standardised techniques, as measured by a reduction in the variation of lymph-node harvest, and, second, improve postoperative outcomes as shown by a reduction in in-hospital mortality and locoregional cancer recurrence.

Data collection

Search strategy and selection criteria

We did a systematic literature search of the Medline, Embase, Web of Science, and Cochrane Library 2014 (issue 5) databases for articles published between Jan 1, 1990, and June 30, 2014 containing the terms "(o)esophagectomy", "gastrectomy", "(o)esophageal cancer", "gastric cancer", "surgery", "laparoscopy", "thoracoscopy", and the medical subject headings "(o) esophagectomy", "gastrectomy", "stomach neoplasms", "(o)esophageal neoplasms", "surgical procedures", "operative", "evidence-based medicine", "evidence-based surgery", and "evidence-based practice". The electronic

For the **Current Controlled Trials Register** see http://www.controlled-trials.com

search was supplemented by a manual search of published abstracts from relevant conference proceedings (2010–13). Reference lists of all relevant studies and the search included the Current Controlled Trials Register.

Two independent observers (SRM and TW) scrutinised all the citation abstracts identified by the search to determine eligibility for inclusion in this study. Publications in any language were included if they met all the following criteria (figure): the study was a multicentre RCT; the study only focused on patients undergoing treatment for oesophageal or gastric cancer; the study had to have been published after 1990 (to ensure that the studies included showed the present surgical and perioperative management strategies); and the study had to include at least one group in which the patients underwent surgery alone without chemotherapy or radiotherapy. For these studies, only the surgery-alone group was included in the analysis to avoid the introduction of an additional confounding variable with the use of chemotherapy and radiotherapy. For RCTs with patients randomly assigned to various surgical techniques, both groups were included in the analysis. The study also had to report either an average lymph-node harvest with SD or range, or the proportion of patients with a locoregional cancer recurrence with a median of at least 5 years' follow-up or quotation of 5 year survival to be eligible for inclusion.

Outcomes and analysis

The outcome measures assessed in this Review were variation in lymph-node harvest, percentage of adjusted in-hospital mortality and locoregional cancer recurrence. The coefficient of variation for lymph-node harvest was used to assess variation in lymph-node resection. We calculated the coefficient of variation by dividing the SD by the mean. Adjusted percentage in-hospital mortality and locoregional cancer recurrence were calculated for each study by division of the absolute percentage values of these outcomes by their respective median values for all studies included in the analysis.

A ten-point checklist was used to assess the design and quality of surgery undertaken in each RCT (panel). This checklist was broadly divided into three main areas: standardisation of surgical technique, credentialing of surgeons, and monitoring of performance during the trial.[7] Each RCT was assessed by two independent investigators (SRM and TW), and then discussed to ensure agreement. The study quality of RCTs was also assessed with the Jadad criteria.[2] Regression modelling was used to identify the most important factors in surgical study design that are associated with the dependent variables of variation in lymph-node harvest (coefficient of variation), percentage in-hospital mortality and locoregional cancer recurrence. The independent variables that were included in the linear regression analyses included the number of patients and centres, the patient-to-centre ratio, the country (east Asian *vs* European and North American countries and Australia), the Jadad score, and study start date. Education, surgical standardisation, and standardisation of the extent of lymphadenectomy were variables included within the domain of standardising surgical technique, and case volume, operative reports, assessment videos, and live operating room assessments were variables included in the credentialing of surgeons. Monitoring of performance

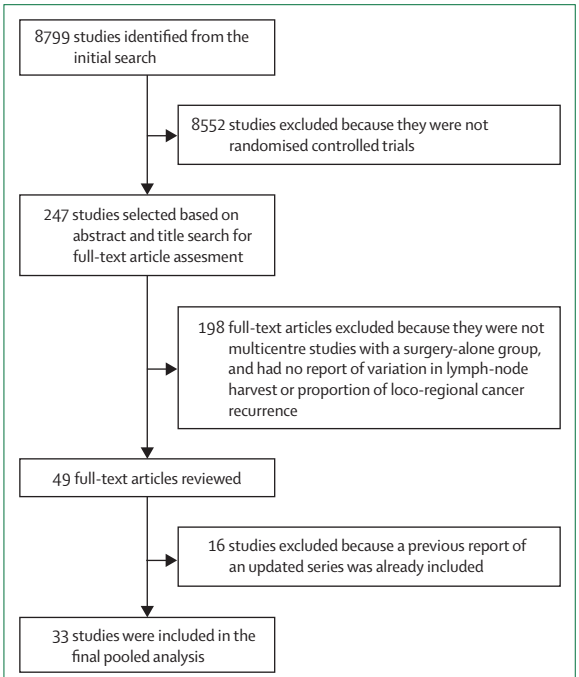

Figure: **Systematic search and selection strategy**

Panel: **Ten-point checklist used to assess the study design for each surgical randomised controlled trial included**

Identification of the importance of standardisation of surgical techniques:

1 Pretrial education through written information, videos, or demonstration
2 Standardisation of surgical approach
3 Standardisation of the extent of lymphadenectomy

Method used to credential surgeons:

4 Case or procedural volume
5 Operative reports
6 Video assessment
7 Live operating room assessment

Method of monitoring performance during the randomised controlled trial:

8 Video submission and assessment
9 Monitoring of data including clinical outcome measures
10 Centralised assessment of pathological changes

	Recruitment	Location	Oesophageal or gastric cancer	Study focus	Number of surgery patients	Centres	Patient-to centre-ratio	Jadad score[2]
Ando[8]	1993–97	Japan	Oesophageal	Adjuvant chemotherapy	122	17	7·18	2
Ando[9]	1988–91	Japan	Oesophageal	Adjuvant chemotherapy	100	12	8·33	2
Bajetta[10]	1992–97	Italy	Gastric	Adjuvant chemotherapy	136	32	4·25	3
Bang[11]	2006–09	China, South Korea, Taiwan	Gastric	Adjuvant chemotherapy	515	37	13·92	2
Bouche[12]	1989–97	France	Gastric	Adjuvant chemotherapy	133	64	2·08	2
Di Constanzo[13]	1995–2000	Italy	Gastric	Adjuvant chemotherapy	128	33	3·88	3
Macdonald[14]	1991–98	USA	Gastric/GOJ	Adjuvant chemoradiotherapy	275	9	30·56	1
Miyashiro[15]	1993–98	Japan	Gastric	Adjuvant chemotherapy	133	11	12·09	3
Nakajima[16]	1988–92	Japan	Gastric	Adjuvant chemotherapy	285	7	40·71	3
Nakajima[17]	1997–2001	Japan	Gastric	Adjuvant chemotherapy	95	7	13·57	2
Nashimoto[18]	1993–94	Japan	Gastric	Adjuvant chemotherapy	123	13	9·46	2
Nitti (FAMTX)[19]	1991–98	Europe	Gastric/GOJ	Adjuvant chemotherapy	103	23	4·48	2
Nitti (FEMTX)[19]	1990–98	International	Gastric	Adjuvant chemotherapy	100	16	6·25	2
Sasako[20]	2001–04	Japan	Gastric	Adjuvant chemotherapy	519	109	4·76	3
Teoh/Chiu[21,22]	2000–04	Hong Kong	Oesophageal	Definitive chemoradiotherapy *vs* surgery	44	5	8·8	2
Cunningham[23]	1994–2002	UK	Gastro-oesophageal	Perioperative chemotherapy	253	50	5·06	3
Allum[24]	1992–98	UK	Oesophageal	Neaodjuvant chemotherapy	402	42	9·57	3
Boonstra[25]	1989–96	Netherlands	Oesophageal	Neoadjuvant chemotherapy	84	6	14	3
Burmeister[26]	1994–2000	Australia	Oesophageal	Neoadjuvant chemoradiotherapy	128	25	5·12	3
Kelsen[27]	1990–95	USA	Oesophageal	Neoadjuvant chemotherapy	129	123	1·05	3
Schuhmacher[28]	1999–2004	Germany	Gastric	Neoadjuvant chemotherapy	68	14	4·86	2
Ychou[29]	1995–2003	France	Gastro-oesophageal	Perioperative chemotherapy	111	28	3·96	2
Tepper[30]	1997–2000	USA	Oesophageal	Neoadjuvant chemoradiotherapy	26	18	1·44	1
van Hagen[31]	2004–08	Netherlands	Oesophageal	Neoadjuvant chemoradiotherapy	188	6	31·33	3
Bonenkamp_D1[32]	1989–93	Netherlands	Gastric	D1 *vs* D2 lymphadenectomy	380	80	4·75	2
Bonenkamp_D2[32]	1989–93	Netherlands	Gastric	D1 *vs* D2 lymphadenectomy	331	80	4·14	2
Sasako_D2[33]	1995–2001	Japan	Gastric	D2 *vs* D2 with para-aortic nodal dissection	263	24	10·96	3
Sasako_D2PAN[33]	1995–2001	Japan	Gastric	D2 *vs* D2 with para-aortic nodal dissection	260	24	10·83	3
Degiuli_D1[34]	1999–2002	Italy	Gastric	D1 *vs* D2 lymphadenectomy	133	5	26·6	3
Degiuli_D2[34]	1999–2002	Italy	Gastric	D1 *vs* D2 lymphadenectomy	134	5	26·8	3
Yonemura_D2[35]	1995–2002	Japan, South Korea & Taiwan	Gastric	D2 *vs* D4 lymphadenectomy	135	13	10·38	1
Yonemura_D4[35]	1995–2002	Japan, South Korea & Taiwan	Gastric	D2 *vs* D4 lymphadenectomy	134	13	10·31	1
Kulig_D2[36]	1999–2003	Poland	Gastric	D2 *vs* D2+ lymphadenectomy	141	10	14·1	3
Kulig_D2+[36]	1999–2003	Poland	Gastric	D2 *vs* D2+ lymphadenectomy	134	10	13·4	3
Imamura_Burs[37]	2002–07	Japan	Gastric	Bursectomy *vs* non-bursectomy	104	11	9·45	3
Imamura_non Burs[37]	2002–07	Japan	Gastric	Bursectomy *vs* non-bursectomy	106	11	9·64	3
Omloo_TH[38]	1994–2000	Netherlands	Oesophageal	Transhiatal *vs* transthoracic oesophagectomy	95	2	47·5	3
Omloo_TT[38]	1994–2000	Netherlands	Oesophageal	Transhiatal *vs* transthoracic oesophagectomy	110	2	55	3
Sasako_LTA[39]	1995–2003	Japan	Gastric	Left thoracoabdominal *vs* transhiatal	85	27	3·15	3

(Table 1 continues on next page)

	Recruitment	Location	Oesophageal or gastric cancer	Study focus	Number of surgery patients	Centres	Patient-to centre-ratio	Jadad score[2]
(Continued from previous page)								
Sasako_TH[39]	1995–2003	Japan	Gastric	Left thoracoabdominal *vs* transhiatal	82	27	3·04	3
Biere_Open[40]	2009–11	Netherlands	Oesophageal	Minimally invasive *vs* open oesophagectomy	56	5	11·2	3
Biere_MIO[40]	2009–11	Netherlands	Oesophageal	Minimally invasive *vs* open oesophagectomy	59	5	11·8	3

GOJ=gastro-oesophageal junction. FAMTX=fluorouracil, doxorubicin, and methotrexate chemotherapy. FEMTX=fluorouracil, epirubicin, and methotrexate. Burs=bursectomy.TH=transhiatal. TT=transthoracic. LTA=left thoracoabdominal. MIO=minimally invasive oesophagectomy.

Table 1: **Randomised controlled trials with a description of the focus of the study and the number of participants undergoing surgery alone**

included the variables video assessments, data monitoring, and pathology. Surgical standardisation refers to the steps needed to maintain surgical approaches that do not differ between oesophageal or gastric resections (eg, left thoracoabdominal, transhiatal, Ivor Lewis or cervical or thoracic anastomosis, the use of drains, etc). Extent of lymphadenectomy refers to a similar approach used for a lymphadenectomy as part of the surgical procedure (eg, D1 or D2 lymphadenectomy, or two-field or three-field lymphadenectomy). A negative β coefficient suggests that the independent variable is associated with low variability in lymph-node harvest, and a reduced percentage of in-hospital mortality and locoregional cancer recurrence in the regression model.

Findings

Selected studies

From the scientific literature, 33 RCTs[8–40] consisting of 42 surgery-alone groups met the inclusion criteria and were included in this Review (figure). 20 studies[10–20,28,32–37,39] mainly focused on the treatment of gastric cancer, 11 on the treatment of oesophageal cancer;[8,9,21,22,24–27,30,31,38,40] and two included both gastric and oesophageal cancer.[23,29] 24 studies[8–31] compared surgery alone versus surgery combined with neoadjuvant or adjuvant chemotherapy or radiotherapy, and five studies[32–36] investigated the extent of lymphadenectomy (D level) during gastrectomy. Other studies compared bursectomy versus non-bursectomy gastrectomy,[37] transhiatal versus trans-thoracic oesophagectomy,[38] transhiatal versus left thoracoabdominal oesophagectomy,[39] and minimal invasive versus open oesophagectomy[40] (table 1). 7045 patients were included from 1086 centres, and a median of 9·5 (range 1·1–55) resections were done in each centre. The median Jadad score for the studies included was 3 (range 1–3); most studies lost points from the Jadad score because their design did not include masking.

Assessment of surgical quality

The assessment of surgical quality by the ten-point checklist is described in table 2 for each RCT included in this Review. 19 (45%) of the 42 surgical groups had some form of pretrial education (visits, videos, or manuals) about the standard of surgery needed for participation in the investigation. 34 (81%) surgical groups attempted to standardise surgical techniques, and 29 (69%) attempted to standardise the extent of lymphadenectomy. Surgical case volume (17 of 42; 41% was the most widely used method to credential the participating surgeons before entry to the study. Other methods included assessment of operative reports (six of 42; 14%), live operating-room assessment (six of 42; 14%), and video assessment (one of 42; 2%). The most common methods of monitoring of surgical performance during the trial period were data monitoring (20 of 42; 48%), assessment of pathological changes (14 of 42; 33%), and video assessment (eight of 42; 19%).

Variation in lymph-node harvest

3917 patients were included in this analysis, with a median lymph-node harvest of 30·5 (range 12·2–103·3) for the six oesophagectomy groups, and 34·8 (12·1–103·3) for the 23 gastrectomy groups that measured this outcome (table 3). The coefficient of variation for lymph-node harvest also varied between the surgical groups with a median of 0·30 (range 0·10–0·56) for the oesophagectomy studies and 0·15 (range 0·07–0·61) for the gastrectomy studies. Regression modelling (table 4) showed that the number of centres (β=0·01; p=0·04) and the patient-to-centre ratio (β=0·01; p=0·01) were associated with a slight rise in the coefficient of variation for lymph-node harvest. An increased Jadad score (β=–0·18; p=0·01), credentialing of a surgeon through operative reports (β=–0·18; p=0·04), and assessment of performance through data monitoring (β=–0·2; p=0·03) were associated with a reduced coefficient of variation for lymph-node harvest.

In-hospital mortality

Median in-hospital mortality for the 25 of 42 surgical groups that reported this outcome (3807 patients) was 3% (range 0–10). The results of regression modelling (table 4) showed that credentialing of surgeons by case

	Education	Surgical standardisation	Extent of lymphadenectomy	Case volume	Operative reports	Videos	Operating room evaluation	Videos	Data monitoring	Pathology review
Ando[8]	N	Y	N	N	N	N	N	N	Y	N
Ando[9]	N	Y	N	N	N	N	N	N	N	N
Bajetta[10]	N	Y	Y	N	Y	N	N	N	Y	Y
Bang[11]	Y	Y	Y	Y	N	N	N	N	N	N
Bouche[12]	N	Y	Y	N	N	N	N	N	N	N
Di Constanzo[13]	N	N	N	N	N	N	N	N	N	N
Macdonald[14]	N	N	N	N	N	N	N	N	Y	Y
Miyashiro[15]	N	Y	Y	N	N	N	N	N	N	N
Nakajima[16]	N	N	N	N	N	N	N	N	Y	N
Nakajima[17]	N	N	N	N	N	N	N	N	Y	N
Nashimoto[18]	N	N	N	N	N	N	N	N	N	N
Nitti (FAMTX)[19]	N	Y	Y	N	N	N	N	N	Y	N
Nitti (FEMTX)[19]	N	Y	N	N	N	N	N	N	Y	N
Sasako[20]	N	Y	Y	N	N	N	N	N	N	N
Teoh/Chiu[21,22]	N	Y	Y	N	N	N	N	N	N	N
Cunningham[23]	N	Y	N	N	N	N	N	N	Y	N
Allum[24]	N	N	N	N	N	N	N	N	N	N
Boonstra[25]	Y	Y	Y	N	N	N	N	N	N	N
Burmeister[26]	N	N	N	N	N	N	N	N	Y	N
Kelsen[27]	Y	N	N	N	N	N	N	N	N	N
Schuhmacher[28]	N	Y	Y	N	Y	N	N	N	Y	Y
Ychou[29]	N	Y	Y	N	N	N	N	N	N	N
Tepper[30]	N	Y	N	N	N	N	N	N	N	N
van Hagen[31]	N	Y	Y	N	N	N	N	N	N	Y
Bonenkamp_D1[32]	Y	Y	Y	N	N	N	Y	N	Y	Y
Bonenkamp_D2[32]	Y	Y	Y	N	N	N	Y	N	Y	Y
Sasako_D2[33]	Y	Y	Y	Y	N	N	N	Y	Y	Y
Sasako_D2PAN[33]	Y	Y	Y	Y	N	N	N	Y	Y	Y
Degiuli_D1[34]	Y	Y	Y	Y	Y	N	N	N	Y	Y
Degiuli_D2[34]	Y	Y	Y	Y	Y	N	N	N	Y	Y
Yonemura_D2[35]	Y	Y	Y	Y	N	N	N	Y	N	N
Yonemura_D4[35]	Y	Y	Y	Y	N	N	N	Y	N	N
Kulig_D2[36]	Y	Y	Y	Y	N	N	N	N	Y	Y
Kulig_D2+[36]	Y	Y	Y	Y	N	N	N	N	Y	Y
Imamura_Burs[37]	Y	Y	Y	Y	N	N	N	N	N	N
Imamura_non Burs[37]	Y	Y	Y	Y	N	N	N	N	N	N
Omloo_TH[38]	N	Y	Y	Y	N	N	Y	N	N	Y
Omloo_TT[38]	N	Y	Y	Y	N	N	Y	N	N	Y
Sasako_LTA[39]	Y	Y	Y	Y	Y	N	N	Y	Y	N
Sasako_TH[39]	Y	Y	Y	Y	Y	N	N	Y	Y	N
Biere_Open[40]	Y	Y	Y	Y	N	N	Y	Y	N	N
Biere_MIO[40]	Y	Y	Y	Y	N	Y	Y	Y	N	N

Y=yes. N=no. FAMTX=fluorouracil, doxorubicin, and methotrexate chemotherapy. FEMTX=fluorouracil, epirubicin, and methotrexate. Burs=bursectomy.TH=transhiatal. TT=transthoracic. LTA=left thoracoabdominal. MIO=minimally invasive oesophagectomy.

Table 2: **Assessment of surgery design in randomised controlled trials (see panel)**

volume (β=–3·31; p=0·04), assessment of operative reports (β=–2·32; p=0·03), and standardisation of surgical techniques (β=–7·97; p=0·01) were associated with a reduction in percentage-adjusted in-hospital mortality.

Locoregional cancer recurrence

Median locoregional cancer recurrence for 28 of the 42 surgical groups (5048 patients) was 12·9% (range 1·1–46·2). The appendix describes the definition and reporting of locoregional cancer recurrence by the

See Online for appendix

	Lymph-node harvest (range orSD)	Coefficient of variation in lymph-node harvest	Perioperative mortality (%)*	Locoregional recurrence (%)
Ando[8]	..	..	..	30 (30%)
Ando[9]	..	..	..	56 (45·9%)
Bajetta[10]	25 (2–87)	0·13	..	23 (18%)
Bang[11]	43·6 (16·7)	0·38	..	..
Bouche[12]	17·5 (1·2)	0·57	..	15 (11·3%)
Di Constanzo[13]	18 (2–68)	0·15	..	20 (16%)
Macdonald[14]	..	..	..	127 (46·2%)
Miyashiro[15]†	..	..	..	4 (3%)
Nakajima[16]†	..	..	..	3 (1·1%)
Nakajima[17]†	..	..	..	4 (4·2%)
Nashimoto[18]†	12·12 (6·7)	0·55	..	2 (1·6%)
Nitti (FAMTX)[19]†	23 (4–74)	0·14	..	12 (11·7%)
Nitti (FEMTX)[19]†	13·5 (0 –47)	0·16	..	3 (3%)
Sasako[20]	..	..	..	71 (13·7%)
Teoh/Chiu[21,22]	12·2 (6·7)	0·55	3 (6·8%)	14 (31·8%)
Cunningham[23]†	..	..	15 (5·9%)	52 (20·6%)
Allum[24]†	..	..	40 (10%)	49 (12·2%)
Boonstra[25]	..	..	3 (4%)	21 (25%)
Burmeister[26]	..	..	..	19 (14·8%)
Kelsen[27]	..	..	..	27 (20·9%)
Schuhmacher[28]	33 (10–88)	0·11	1 (1·5%)	..
Ychou[29]	19 (2–82)	0·15	..	9 (8·1%)
Tepper[30]†	..	..	1 (2·1%)	3 (11·5%)
Van Hagen[31]	18 (IQR 12·5 –27)	0·10	8 (4%)	17 (9·3%)
Bonenkamp_D1[32]	18·4 (0 –73)	0·61	15 (4%)	56 (14·7%)
Bonenkamp_D2[32]	31·5 (0 –106)	0·6	32 (10%)	40 (12·1%)
Sasako_D2[33]	54 (14–161)	0·44	2 (0·8%)	24 (9·1%)
Sasako_D2PAN[33]	74 (30–235)	0·48	2 (0·8%)	23 (8·8%)
Degiuli_D1[34]	28·2 (2–104)	0·18	4 (3%)	..
Degiuli_D2[34]	37·3 (11–124)	0·14	3 (2·2%)	..
Yonemura_D2[35]	42·7 (18·7)	0·09	1 (0·8%)	..
Yonemura_D4[35]	68·7 (33)	0·07	5 (3·9%)	..
Kulig_D2[36]	23 (95% CI 21·2 –24)	0·38	7 (4·9%)	..
Kulig_D2+[36]	28 (95% CI 25·1 –31)	0·61	3 (2·2%)	..
Imamura_Burs[37]	38 (11–98)	0·1	1 (1%)	..
Imamura_non Burs[37]	37 (7–97)	0·11	1 (0·9%)	..
Omloo_TH[38]	16 (9)	0·56	2 (2%)	13 (13·7%)
Omloo_TT[38]	31 (14)	0·45	5 (4%)	16 (14·5%)
Sasako_LTA[39]	68 (14–147)	0·08	3 (4%)	..
Sasako_TH[39]	60 (16–160)	0·08	0 (0%)	..
Biere_Open[40]	21 (7–47)	0·13	1 (2%)	..
Biere_MIO[40]	20 (3–44)	0·15	2 (3%)	..

FAMTX=fluorouracil, doxorubicin, and methotrexate chemotherapy. FEMTX=fluorouracil, epirubicin, and methotrexate. Burs=bursectomy.TH=transhiatal. TT=transthoracic. LTA=left thoracoabdominal. MIO=minimally invasive oesophagectomy. *In-hospital and 30 day mortality grouped together. †Reports of these studies describe local versus distant recurrence, with no definition of local recurrence. No description of regional cancer recurrences—for the purposes of analyses these were treated as locoregional recurrences.

Table 3: **Surgical outcomes from randomised controlled trials**

individual studies included. Credentialing of surgeons through operating room assessments (β=–2·47; p=0·07) and video monitoring of surgical performance (β=–4·20; p=0·07) was associated with a reduction in adjusted locoregional cancer recurrence, although this was not significant (table 4).

Conclusions

The results of this Review suggest a large amount of heterogeneity in study design and surgical-quality assessments in multicentre RCTs for the treatment of gastro-oesophageal cancer. A similar degree of heterogeneity of study design and surgical quality assessment was present in trials from different countries and in trials with a different primary aim (eg, investigation of the extent of lymphadenectomy or the use of chemotherapy or radiotherapy). Additionally, the coefficient of variation for lymph-node harvest (range 0·07–0·61), percentage of in-hospital mortality (0–10%) and locoregional cancer recurrence (1·1–46·2%) in those trials showed wide variation. Important factors that were associated with variation in lymph-node harvest were the number of centres included in the study and the patient-to-centre ratio. Credentialing of surgeons before enrolment in the study through assessment of operative reports and monitoring of performance through data checking were important factors in the study design that reduced the variation in lymph-node harvest in those individual trials. Some factors in the study design that helped to reduce in-hospital mortality included credentialing surgeons through assessment of procedural volume and operative reports, and standardisation of surgical techniques. Credentialing of surgeons through operating-room assessments and monitoring of performance during the trial by video assessment was associated with a non-significant reduction in adjusted locoregional cancer recurrence.

Although the Jadad system is often used in the assessment of surgical RCTs, the present analysis suggests that it is of little use in assessing the quality of surgery because the score did not correlate with in-hospital mortality or locoregional cancer recurrence. This Review identified three areas that provide the basis for surgical quality assurance programmes in multicentre RCTs because of their potential effect on the variability of outcomes. First, pretrial standardisation of surgical techniques can take the form of practical or video demonstrations or written information, dependent on the complexity of surgical interventions and the familiarity of surgeons with the intervention techniques. Second, credentialing of a surgeon can be done through a combined assessment of procedural volume and direct observation of operative performances via live operating-room assessments or unedited video assessments. Trials should only start recruitment in a specific centre when appropriate surgical standards are achieved by most participating surgeons in that centre.[5] Third, monitoring of performance during the trial can be obtained through video or photo assessments of the operative field to ensure adherence to standard techniques. Regular audits

	Coefficient of variance in lymph-node harvest			Adjusted in-hospital mortality			Adjusted locoregional cancer recurrence		
	β	95% CI	p value	β	95% CI	p value	β	95% CI	p value
Number of patients	<0·001	−0·001 to 0·001	0·40	0·01	−0·01 to 0·01	0·40	<0·01	−0·01 to 0·01	0·81
Number of centres	0·01	0·001 to 0·01	0·04	0·01	−0·03 to 0·04	0·72	0·01	−0·02 to 0·02	0·89
Patient-to-centre ratio	0·01	0·003 to 0·02	0·01	−0·01	−0·04 to 0·04	0·88	0·01	−0·03 to 0·04	0·71
Country: east Asian countries *vs* European and North American countries, and Australia	0·03	−0·13 to 0·19	0·68	−0·59	−1·39 to 0·20	0·13	0·01	−0·85 to 0·87	0·98
Jadad score	−0·18	−0·30 to −0·06	0·01	−0·08	−0·64 to 0·47	0·76	−0·45	−1·10 to 0·20	0·16
Study start date	0·01	−0·01 to 0·02	0·59	−0·02	−0·11 to 0·08	0·74	−0·03	−0·13 to 0·08	0·61
SST education	0·01	−0·17 to 0·19	0·90	2·48	−0·42 to 5·37	0·09	1·05	−0·66 to 2·76	0·21
SST surgical	−0·25	−0·77 to 0·28	0·34	−7·97	−13·66 to −2·29	0·01	0·89	−0·27 to 2·05	0·13
SST extent of lymphadenectomy	0·22	−0·24 to 0·67	0·34	−0·31	−3·08 to 2·46	0·81	−0·70	−1·89 to 0·48	0·23
CSE case volume	0·02	−0·15 to 0·18	0·85	−3·31	−6·57 to −0·43	0·04	1·45	−1·44 to 4·34	0·31
CSE operative reports	−0·18	−0·36 to −0·004	0·04	−2·32	−4·41 to −0·22	0·03	−1·03	−3·68 to 1·61	0·42
CSE videos	−0·16	−0·62 to 0·30	0·47	0·46	−1·69 to 2·61	0·64	..	..	..
CSE operative room assessment	0·15	−0·04 to 0·35	0·11	−1·24	−4·05 to 1·56	0·34	−2·47	−5·28 to 0·35	0·07
MP videos	−0·1	−0·23 to 0·12	0·52	0·59	−0·79 to 1·97	0·36	−4·20	−9·18 to 0·78	0·07
MP data monitoring	−0·2	−0·39 to −0·02	0·03	2·50	−0·20 to 5·21	0·10	0·29	−0·76 to 1·35	0·57
MP pathological changes	0·11	−0·07 to 0·28	0·22	..	..	..	1·34	−0·23 to 2·9	0·10

SST=standardisation of surgical technique. CSE=credentialing of surgical experience before enrolment. MP=monitoring of performance during the trial. Adjusted in-hospital mortality=% in-hospital mortality (individual study)/median % in-hospital mortality for all studies. Monitoring of pathology excluded from inhospital mortality assessment because it does not affect this variable. Adjusted locoregional cancer recurrence (LCR) score=% LCR (individual study)/median % LCR (all studies).

Table 4: **Regression analysis for coefficient of variance in lymph-node harvest, adjusted in-hospital mortality, and adjusted locoregional cancer recurrence**

through data monitoring provide feedback to participating surgeons to further standardise surgical approach.[41] The use of pathological assessment of the resected specimen as a measure of surgical quality needs standardised techniques for lymph-node retrieval to eliminate the pathologist's judgment as a potential source of bias.[42] Educational research showed that assessment of operative performance in live operating rooms or via unedited videos should be undertaken with valid and reliable competency assessment methods. The laparoscopic colorectal national training programme in England[43] showed that the assessment of specialists with objective assessment methods is achievable at a national level.

These trials have limitations that need to be considered during interpretation of results from this Review. The protocols of the RCTs were only available in five of the studies included in this study, and might have, therefore, received a lower score than others in the ten-point checklist because of insufficient information. Similarly, the prevalence of video-assessment methods to credential surgeons before enrolment in the trial was low, which suggests that the analysis was underpowered to show a significant difference in this variable. Furthermore, although some surgical trials described methods to standardise surgery and monitor performance, few of the studies reported rates of adherence to the study design protocol with standardisation of surgery and monitoring of performance, and the outcome of these assessments. Additionally, the variation noted in clinical outcomes could have been affected by other confounding variables (eg, age, tumour status, medical comorbidities, postoperative problems, and the overall quality of care including recovery). The inclusion criteria for tumour stages studied varied within the RCTs (eg, in the trial by Tepper and colleagues,[30] T1–3NX including regional thoracic lymph-node [N1] metastases were included), and tumour stages can vary between countries (eg, in Japan, many patients with gastric cancer are diagnosed early because the risk of developing this type of cancer is widely known).[44] Surgery-alone groups were only studied in this review to reduce the effect of other confounding variables, such as standardising chemotherapy or radiotherapy procedures.

The number of centres taking part in each trial was well described; however, the number of surgeons included was described in less than 10% of the trials. Therefore, we were not able to examine the effect that surgeons' skills and experience had on this study. However, recruitment for RCTs is most usually done on a centre-specific rather than a surgeon-specific basis. This investigation provides evidence that credentialing of individual surgeons within centres taking part in RCTs is an important factor that affects clinical outcome.

The median lymph-node yield range was high (30–35) both for patients who had had an oesophagectomy or a gastrectomy—perhaps because 17 Asian studies (41%) were included in this Review, and that the early, large studies[23,24] did not report a lymph-node yield.

Nevertheless, the coefficient of variation in lymph-node harvest for each surgical group was analysed, not the absolute lymph-node yield.

The implementation of and adherence to robust, surgical-standardisation protocols within RCTs needs participating surgeons to commit to a high quality of surgery. Surgeons with a high volume of procedures in specialised units have previously used video assessments to monitor their performance and have been able to attribute credit to this approach.[33,35,39,40] Attempts to standardise surgery that are not routinely done by participating surgeons could introduce another variable to performance as surgeons climb the learning curve for the new technique. Credentialing of surgeons by procedural volume before enrolment in the trial ensures that surgeons have reached an adequate level in their surgical ability. Monitoring of performance allows for the assessment of any evolution in surgical proficiency during the trial. Investigators in oncological trials should be aware that the outcomes might be affected by surgical technique, enforcing the principle that surgery needs to be standardised. However, if standardisation is not possible, skill-based studies could have a role in determining where surgeons are allowed to do the procedure according to their preference and experience.

Multicentre RCTs represent the highest level of evidence for any surgical intervention, and are used as the basis of meta-analyses that often change surgical practice and health-care policy. These analyses often use Cochrane Q or I^2 statistics to quantify heterogeneity and Egger tests to establish any bias between the studies. However, these statistical tests do not reliably measure the variation within and between studies, which has been described in this Review. Therefore, the variation in clinical outcomes within and between multicentre RCTs, and the strength of their previous recommendations, should be viewed with caution. This issue is underlined in a communication by the principal investigator of the Medical Research Council D1 versus D2 gastrectomy RCT:[45] "The surgical community needs to address quality issues required for optimum surgical performance in oncological randomised controlled trials".

In conclusion, credentialing of surgeons before enrolment in a study, standardisation of surgical techniques, and monitoring of surgical performances during a trial have a positive effect on the quality of RCTs. Future studies should include methods to assure surgical quality, to reduce variation in clinical outcome and improve the reliability of trial findings.

Contributors

GBH conceived the idea for the study. GBH, SRM, MN, EWS, JJBvL, and MS designed the study, and analysed or interpreted the data. SRM and TW searched the scientific literature. SRM obtained the data and produced the figure and the panel. GBH, SRM, TW, EWS, JJBvL, and MS wrote the text.

Declaration of interests

We declare no competing interests.

References

1 Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; **61:** 69–90.

2 Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; **17:** 1–12.

3 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273:** 408–12.

4 Higgins JP, Altman DG, Gøtzsche PC, et al, and the Cochrane Bias Methods Group, and the Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; **343:** d5928.

5 Bunt AM, Hermans J, Boon MC, et al. Evaluation of the extent of lymphadenectomy in a randomized trial of Western- versus Japanese-type surgery in gastric cancer. *J Clin Oncol* 1994; **12:** 417–22.

6 Macdonald JS, Benedetti N, Estes N, et al. Postoperative combined radiation and chemotherapy improves disease-free survival (DFS) and overall survival (OS) in resected adenocarcinoma of the stomach and gastroesophageal junction: update of the results of intergroup study INT-0116 (SWOG 9008). Paper Presented at: 2004 American Society of Clinical Oncology Gastrointestinal Cancers Symposium; January 22–24, 2004; San Francisco, CA, USA.

7 Foster JD, Mackenzie H, Nelson H, Hanna GB, Francis NK. Methods of quality assurance in multicentre trials in laparoscopic colorectal surgery: a systematic review. *Ann Surg* 2014; **260:** 220–29.

8 Ando N, Iizuka T, Ide H, et al, and the Japan Clinical Oncology Group. Surgery plus chemotherapy compared with surgery alone for localized squamous cell carcinoma of the thoracic esophagus: a Japan Clinical Oncology Group Study--JCOG9204. *J Clin Oncol* 2003; **21:** 4592–6.

9 Ando N, Iizuka T, Kakegawa T, et al. A randomized trial of surgery with and without chemotherapy for localized squamous carcinoma of the thoracic esophagus: the Japan Clinical Oncology Group Study. *J Thorac Cardiovasc Surg* 1997; **114:** 205–9.

10 Bajetta E, Buzzoni R, Mariani L, et al. Adjuvant chemotherapy in gastric cancer: 5-year results of a randomised study by the Italian Trials in Medical Oncology (ITMO) Group. *Ann Oncol* 2002; **13:** 299–307.

11 Bang YJ, Kim YW, Yang HK, et al, and the CLASSIC trial investigators. Adjuvant capecitabine and oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): a phase 3 open-label, randomised controlled trial. *Lancet* 2012; **379:** 315–21.

12 Bouché O, Ychou M, Burtin P, et al, and the Fédération Francophone de Cancérologie Digestive Group. Adjuvant chemotherapy with 5-fluorouracil and cisplatin compared with surgery alone for gastric cancer: 7-year results of the FFCD randomized phase III trial (8801). *Ann Oncol* 2005; **16:** 1488–97.

13 Di Costanzo F, Gasperoni S, Manzione L, et al, and the Italian Oncology Group for Cancer Research. Adjuvant chemotherapy in completely resected gastric cancer: a randomized phase III trial conducted by GOIRC. *J Natl Cancer Inst* 2008; **100:** 388–98.

14 Macdonald JS, Smalley SR, Benedetti J, et al. Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. *N Engl J Med* 2001; **345:** 725–30.

15 Miyashiro I, Furukawa H, Sasako M, et al, and the Gastric Cancer Surgical Study Group in the Japan Clinical Oncology Group. Randomized clinical trial of adjuvant chemotherapy with intraperitoneal and intravenous cisplatin followed by oral fluorouracil (UFT) in serosa-positive gastric cancer versus curative resection alone: final results of the Japan Clinical Oncology Group trial JCOG9206-2. *Gastric Cancer* 2011; **14:** 212–18.

16 Nakajima T, Nashimoto A, Kitamura M, et al, and the Gastric Cancer Surgical Study Group. Adjuvant mitomycin and fluorouracil followed by oral uracil plus tegafur in serosa-negative gastric cancer: a randomised trial. *Lancet* 1999; **354:** 273–77.

17 Nakajima T, Kinoshita T, Nashimoto A, et al, and the National Surgical Adjuvant Study of Gastric Cancer Group. Randomized controlled trial of adjuvant uracil-tegafur versus surgery alone for serosa-negative, locally advanced gastric cancer. *Br J Surg* 2007; **94:** 1468–76.

18 Nashimoto A, Nakajima T, Furukawa H, et al, and the Gastric Cancer Surgical Study Group, Japan Clinical Oncology Group. Randomized trial of adjuvant chemotherapy with mitomycin, fluorouracil, and cytosine arabinoside followed by oral fluorouracil in serosa-negative gastric cancer: Japan Clinical Oncology Group 9206-1. *J Clin Oncol* 2003; **21:** 2282–87.

19 Nitti D, Wils J, Dos Santos JG, et al, and the EORTC GI Group, and the ICCG. Randomized phase III trials of adjuvant FAMTX or FEMTX compared with surgery alone in resected gastric cancer. A combined analysis of the EORTC GI Group and the ICCG. *Ann Oncol* 2006; **17:** 262–69.

20 Sasako M, Sakuramoto S, Katai H, et al. Five-year outcomes of a randomized phase III trial comparing adjuvant chemotherapy with S-1 versus surgery alone in stage II or III gastric cancer. *J Clin Oncol* 2011; **29:** 4387–93.

21 Teoh AYB, Chiu PWY, Yeung WK, Liu SY, Wong SK, Ng EK. Long-term survival outcomes after definitive chemoradiation versus surgery in patients with resectable squamous carcinoma of the esophagus: results from a randomized controlled trial. *Ann Oncol* 2013; **24:** 165–71.

22 Chiu PWY, Chan ACW, Leung SF, et al. Multicenter prospective randomized trial comparing standard esophagectomy with chemoradiotherapy for treatment of squamous esophageal cancer: early results from the Chinese University Research Group for Esophageal Cancer (CURE). *J Gastrointest Surg* 2005; **9:** 794–802.

23 Cunningham D, Allum WH, Stenning SP, et al, and the MAGIC Trial Participants. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med* 2006; **355:** 11–20.

24 Allum WH, Stenning SP, Bancewicz J, Clark PI, Langley RE. Long-term results of a randomized trial of surgery with or without preoperative chemotherapy in esophageal cancer. *J Clin Oncol* 2009; **27:** 5062–67.

25 Boonstra JJ, Kok TC, Wijnhoven BP, et al. Chemotherapy followed by surgery versus surgery alone in patients with resectable oesophageal squamous cell carcinoma: long-term results of a randomized controlled trial. *BMC Cancer* 2011; **11:** 181.

26 Burmeister BH, Smithers BM, Gebski V, et al, and the Trans-Tasman Radiation Oncology Group, and the Australasian Gastro-Intestinal Trials Group. Surgery alone versus chemoradiotherapy followed by surgery for resectable cancer of the oesophagus: a randomised controlled phase III trial. *Lancet Oncol* 2005; **6:** 659–68.

27 Kelsen DP, Winter KA, Gunderson LL, et al, and the Radiation Therapy Oncology Group, and the USA Intergroup. Long-term results of RTOG trial 8911 (USA Intergroup 113): a random assignment trial comparison of chemotherapy followed by surgery compared with surgery alone for esophageal cancer. *J Clin Oncol* 2007; **25:** 3719–25.

28 Schuhmacher C, Gretschel S, Lordick F, et al. Neoadjuvant chemotherapy compared with surgery alone for locally advanced cancer of the stomach and cardia: European Organisation for Research and Treatment of Cancer randomized trial 40954. *J Clin Oncol* 2010; **28:** 5210–18.

29 Ychou M, Boige V, Pignon JP, et al. Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFCD multicenter phase III trial. *J Clin Oncol* 2011; **29:** 1715–21.

30 Tepper J, Krasna MJ, Niedzwiecki D, et al. Phase III trial of trimodality therapy with cisplatin, fluorouracil, radiotherapy, and surgery compared with surgery alone for esophageal cancer: CALGB 9781. *J Clin Oncol* 2008; **26:** 1086–92.

31 van Hagen P, Hulshof MCCM, van Lanschot JJB, et al, and the CROSS Group. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med* 2012; **366:** 2074–84.

32 Bonenkamp JJ, Hermans J, Sasako M, et al, and the Dutch Gastric Cancer Group. Extended lymph-node dissection for gastric cancer. *N Engl J Med* 1999; **340:** 908–14.

33 Sasako M, Sano T, Yamamoto S, et al, and the Japan Clinical Oncology Group. D2 lymphadenectomy alone or with para-aortic nodal dissection for gastric cancer. *N Engl J Med* 2008; **359:** 453–62.

34 Degiuli M, Sasako M, Ponti A, et al, and the Italian Gastric Cancer Study Group. Randomized clinical trial comparing survival after D1 or D2 gastrectomy for gastric cancer. *Br J Surg* 2014; **101:** 23–31.

35 Yonemura Y, Wu CC, Fukushima N, et al, and the East Asia Surgical Oncology Group. Randomized clinical trial of D2 and extended para-aortic lymphadenectomy in patients with gastric cancer. *Int J Clin Oncol* 2008; **13:** 132–37.

36 Kulig J, Popiela T, Kolodziejczyk P, Sierzega M, Szczepanik A, and the Polish Gastric Cancer Study Group. Standard D2 versus extended D2 (D2+) lymphadenectomy for gastric cancer: an interim safety analysis of a multicenter, randomized, clinical trial. *Am J Surg* 2007; **193:** 10–15.

37 Imamura H, Kurokawa Y, Kawada J, et al. Influence of bursectomy on operative morbidity and mortality after radical gastrectomy for gastric cancer: results of a randomized controlled trial. *World J Surg* 2011; **35:** 625–30.

38 Omloo JMT, Lagarde SM, Hulscher JBF, et al. Extended transthoracic resection compared with limited transhiatal resection for adenocarcinoma of the mid/distal esophagus: five-year survival of a randomized clinical trial. *Ann Surg* 2007; **246:** 992–1000.

39 Sasako M, Sano T, Yamamoto S, et al, and the Japan Clinical Oncology Group (JCOG9502). Left thoracoabdominal approach versus abdominal-transhiatal approach for gastric cancer of the cardia or subcardia: a randomised controlled trial. *Lancet Oncol* 2006; **7:** 644–51.

40 Biere SSAY, van Berge Henegouwen MI, Maas KW, et al. Minimally invasive versus open oesophagectomy for patients with oesophageal cancer: a multicentre, open-label, randomised controlled trial. *Lancet* 2012; **379:** 1887–92.

41 Sasako M, Kurokawa Y. Challenges in performing surgical randomized controlled trials in Japan. *Surgery* 2009; **145:** 598–602.

42 Hanna GB, Amygdalos I, Ni M, et al. Improving the standard of lymph node retrieval after gastric cancer surgery. *Histopathology* 2013; **63:** 316–24.

43 Miskovic D, Ni M, Wyles SM, et al, and the National Training Programme in Laparoscopic Colorectal Surgery in England. Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England. *Ann Surg* 2013; **257:** 476–82.

44 Suzuki H, Gotoda T, Sasako M, Saito D. Detection of early gastric cancer: misunderstanding the role of mass screening. *Gastric Cancer* 2006; **9:** 315–19.

45 Cuschieri SA, Hanna GB. Meta-analysis of D1 versus D2 gastrectomy for gastric adenocarcinoma: let us move on to another era. *Ann Surg* 2014; **259:** e90.