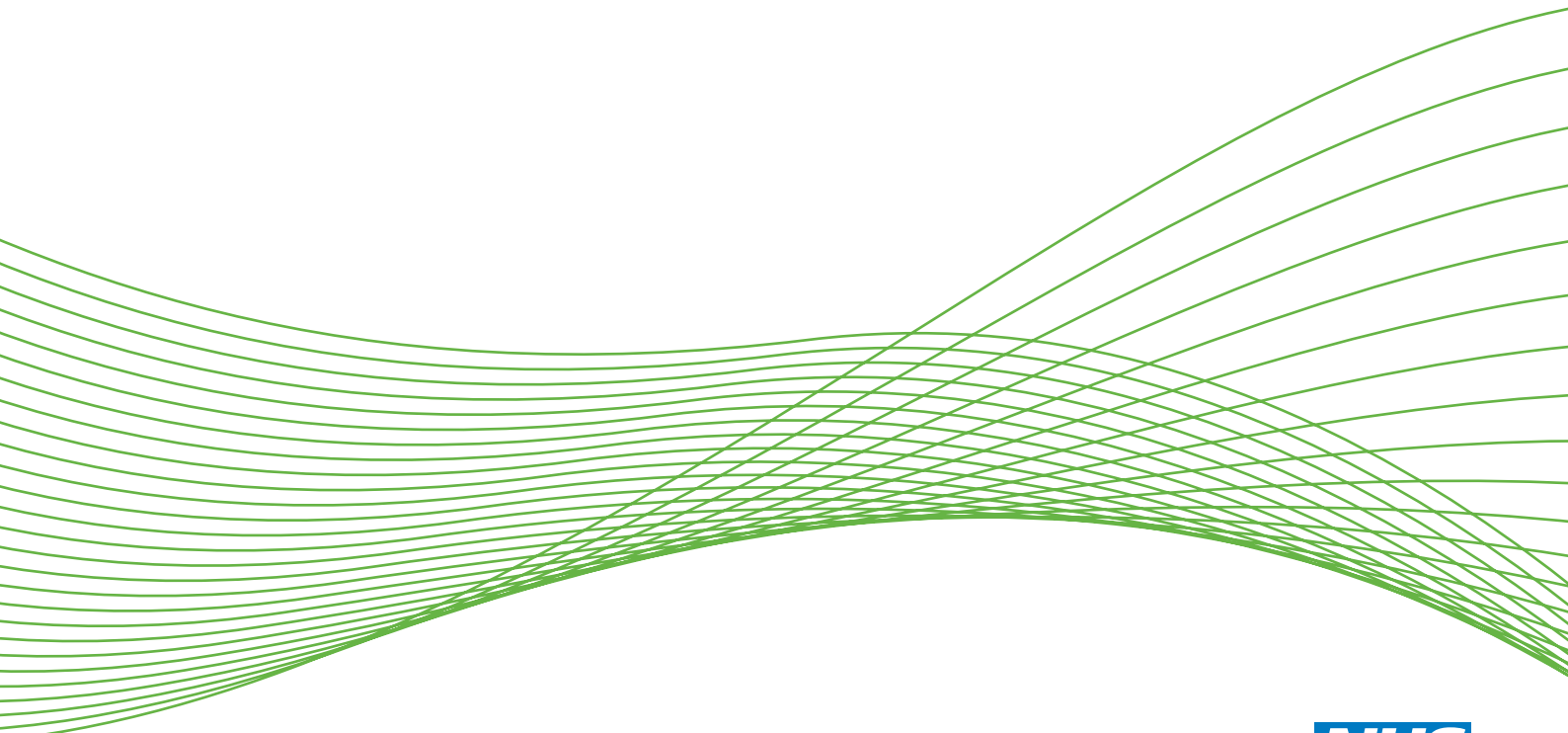# Cancer of Oesophagus or Gastricus – New Assessment of Technology of Endosonography (COGNATE): report of pragmatic randomised trial

IT Russell, RT Edwards, AE Gliddon, DK Ingledew, D Russell, R Whitaker, ST Yeo, SE Attwood, H Barr, S Nanthakumaran and KGM Park

NHS

National Institute for Health Research

Cancer of Oesophagus or Gastricus – New Assessment of Technology of Endosonography (COGNATE): report of pragmatic randomised trial

IT Russell,[1,2]* RT Edwards,[3] AE Gliddon,[1] DK Ingledew,[4] D Russell,[1,2] R Whitaker,[1] ST Yeo,[3] SE Attwood,[5] H Barr,[6] S Nanthakumaran[7] and KGM Park[7]

[1]North Wales Organisation for Randomised Trials in Health (NWORTH), Bangor University, Bangor, UK
[2]College of Medicine, Swansea University, Swansea, UK
[3]Centre for Economics and Policy in Health, Bangor University, Bangor, UK
[4]School of Psychology, Bangor University, Bangor, UK
[5]North Tyneside General Hospital, North Shields, UK
[6]Gloucestershire Royal Hospital, Gloucester, UK
[7]Aberdeen Royal Infirmary, Aberdeen, UK

*Corresponding author

Dedication: We dedicate this report to the memory of Ceri Margaret Bray (1957–2008), first trial manager, whose energy and dedication were crucial to COGNATE's success.

Health Technology Assessment

> ## Criteria for inclusion in the *Health Technology Assessment* journal
> Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.
>
> Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

Abstract

Cancer of Oesophagus or Gastricus – New Assessment of Technology of Endosonography (COGNATE): report of pragmatic randomised trial

IT Russell,[1,2]* RT Edwards,[3] AE Gliddon,[1] DK Ingledew,[4]
D Russell,[1,2] R Whitaker,[1] ST Yeo,[3] SE Attwood,[5] H Barr,[6]
S Nanthakumaran[7] and KGM Park[7]

[1]North Wales Organisation for Randomised Trials in Health (NWORTH), Bangor University, Bangor, UK
[2]College of Medicine, Swansea University, Swansea, UK
[3]Centre for Economics and Policy in Health, Bangor University, Bangor, UK
[4]School of Psychology, Bangor University, Bangor, UK
[5]North Tyneside General Hospital, North Shields, UK
[6]Gloucestershire Royal Hospital, Gloucester, UK
[7]Aberdeen Royal Infirmary, Aberdeen, UK

*Corresponding author ian.russell@bangor.co.uk

Background: Endoscopic ultrasonography is recommended for staging gastro-oesophageal cancers, but has never been evaluated.

Objective: COGNATE (Cancer of Oesophagus or Gastricus – New Assessment of Technology of Endosonography) therefore aimed to evaluate whether adding 'endoscopic ultrasound' (EUS) to the usual staging algorithm changes treatment, improves (quality-adjusted) survival, and uses resources cost-effectively.

Design: Pragmatic parallel-group trial. Patients with gastro-oesophageal cancer received standard staging algorithms. Multidisciplinary teams chose provisional management plans from endoscopic mucosal resection, immediate surgery, surgery after chemotherapy, or chemotherapy and radiotherapy. We used dynamic randomisation to allocate consenting patients remotely by telephone in equal proportions between EUS and not. Thereafter we recorded changes in management plan, use of health-care resources, and three aspects of participant-reported quality of life: generic [measured by European Quality of Life – 5 Dimensions (EQ-5D)], cancer related [Functional Assessment of Cancer Therapy – General scale (FACT-G)] and condition-specific [FACT – Additional Concerns scale (FACT-AC)]. We followed participants regularly until death or the end of the trial – for between 1 and 4.5 years. We devised a quality assurance programme to maintain standards of endosonographic reporting.

Setting: Eight British hospitals, of which two – one Scottish teaching hospital and one English district general hospital – contributed 80% of participants; we combined the other six for analysis.

Participants: Patients were eligible if they had a diagnosis of gastro-oesophageal cancer, had not started treatment, were free of metastatic disease, were fit for surgery (even if not planned) and had American Society of Anesthesiologists and World Health Organization grades of less than 3.

Interventions: *Intervention group*: standard staging algorithm plus EUS; *control group*: standard staging algorithm.

Main outcome measures: *Primary*: quality-adjusted survival. *Secondary*: survival; health-related quality of life (EQ-5D, FACT-G and FACT-AC scales); changes in management plan; and complete resection rate. Although blinding participants was neither possible nor desirable, those responsible for analysis remained blind until the Trial Steering Committee had reviewed the definitive analysis.

Results: We randomised 223 patients, of whom 213 yielded enough data for primary analysis. EUS improved survival adjusted for generic quality of life with a hazard ratio of 0.705 [95% confidence interval (CI) 0.499 to 0.995], and crude survival with a hazard ratio of 0.706 (95% CI 0.501 to 0.996). The benefits of EUS were significantly greater for those with poor initial quality of life, but did not differ between centres. EUS reduced net use of health-care resources by £2860 (95% 'bootstrapped' CI from –£2200 to £8000). Combining benefits and savings shows that EUS is likely to be cost-effective, with 96% probability of achieving the National Institute for Health and Care Excellence criterion of costing of <£20,000 to gain a QALY. There were no serious adverse reactions attributable to EUS. EUS enhanced the management plan for many participants, increased the proportion of tumours completely resected from 80% (44 out of 55) to 91% (48 out of 53), and improved the survival of those who changed plan; although underpinning the significant differences in outcome, none of these process differences was itself significant.

Conclusion: Endoscopic ultrasound significantly improves (quality-adjusted) survival, has the potential to reduce health-care resource use (not statistically significant) and is probably cost-effective (with 96% probability). We recommend research into the best time to evaluate new technologies.

Trial registration: ISRCTN1444215.

Contents

List of abbreviations

5FU	fluorouracil
ANCOVA	analysis of covariance
ASA	American Society of Anesthesiologists
AUC	area under the curve
AUGIS	Association of Upper Gastro-Intestinal Surgeons of Great Britain and Ireland
CD	compact disc
CEAC	cost-effectiveness acceptability curve
CFI	comparative fit index
CI	confidence interval
COGNATE	Cancer of Oesophagus or Gastricus: New Assessment of Technology of Endosonography
CONSORT	Consolidated Standards Of Reporting Trials
CSRI	Client Service Receipt Inventory
CT	computerised tomography
DMEC	Data Monitoring and Ethics Committee
EMR	endoscopic mucosal resection
EQ-5D	European Quality of Life – 5 Dimensions
EQ-VAS	EuroQol Visual Analogue Scale
EuroQol	European Quality of Life instrument
EUS	endoscopic ultrasound
FACT	Functional Assessment of Cancer Therapy
FACT-AC	Functional Assessment of Cancer Therapy – Additional Concerns
FACT-E	Functional Assessment of Cancer Therapy – Oesophageal

FACT-EG	Functional Assessment of Cancer Therapy – Oesophageal and Gastric
FACT-G	Functional Assessment of Cancer Therapy – General
FACT-Ga	Functional Assessment of Cancer Therapy – Gastric
FNA	fine-needle aspiration
FNAC	fine-needle aspiration cytology
GCP	Good Clinical Practice
HRQoL	health-related quality of life
HTA	Health Technology Assessment
ICER	incremental cost-effectiveness ratio
LREC	Local Research Ethics Committee
MDCT	multi-detector row-computed tomography
MDM	multidisciplinary meeting
MDT	multidisciplinary team
MRC	Medical Research Council
MREC	Multicentre Research Ethics Committee
MRI	magnetic resonance imaging
MVA	Missing Values Analysis
N.A.	not applicable
NCCHTA	National Coordinating Centre for Health Technology Assessment
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMB	net monetary benefit
NOGCA	National Oesophago-Gastric Cancer Audit

NR	not recorded		RR	relative risk
NWORTH	North Wales Organisation for Randomised Trials in Health		SAGOC	Scottish Audit of Gastric and Oesophageal Cancer
PCA	*Prescription Cost Analysis*		SD	standard deviation
PET	positron emission tomography		SOP	standard operating procedure
PET-CT	integrated positron emission tomography and computed tomography		SRMR	standardised root-mean-square residual
PI	principal investigator		TEG	Trial Executive Group
PROM	patient-reported outcome measure		Tis	tumour in situ
QALY	quality-adjusted life-year		TNM	staging of tumour, nodes and metastases
R0	complete resection of tumour		TOI	Trial Outcome Index
RCT	randomised controlled trial		TSC	Trial Steering Committee
RMSEA	root-mean-square error of approximation		VAS	visual analogue scale
			WHO	World Health Organization

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or at the end of the table.

Scientific summary

Background

Each year in the UK, 13,000 patients are diagnosed with gastro-oesophageal cancer. It is the fourth most common cause of cancer death. The general prognosis of patients with gastro-oesophageal cancer is poor: fewer than 10% survive 5 years. However there have been many advances in the treatment of these tumours, so it is important to select the most appropriate management plan for each patient. Endoscopic ultrasonography, introduced in the early 1980s, became common practice in the early 2000s. Mounting an ultrasonic probe on an endoscope can improve staging, and guide management, of gastro-oesophageal tumours. Endoscopic ultrasound (EUS) itself is safe and does not impose greater risks on patients than endoscopy.

However there has been no published or current randomised trial to evaluate whether EUS is effective and cost-effective in the management of gastro-oesophageal cancer. So the National Institute of Health Research Health Technology Assessment (NIHR HTA) programme commissioned the COGNATE (Cancer of Oesophagus or Gastricus – New Assessment of Technology of Endosonography) team to evaluate this. Our philosophy was pragmatic in the sense that, after randomising patients to EUS or not, participating centres followed best practice in staging and managing gastro-oesophageal cancer, but without the constraints of a rigid clinical protocol.

Objectives

As the link between staging and managing gastro-oesophageal cancer was not clear, we did not know whether EUS improved management decisions. To monitor the long-term sequelae of EUS, we collected data on participants' treatment plans, changes to them, subsequent progress and use of health care. Evidence that EUS improves choice of treatment and patient outcome would benefit individual patients and the population of patients with gastro-oesophageal cancer by targeting resources better.

Thus COGNATE evaluated, not the accuracy of EUS, but the effect it has on patient management and subsequent survival and quality of life. Most previous assessments of EUS staging neglected patients with non-surgical treatment but COGNATE adopted a broader approach because, if EUS leads to less surgery, it is as important to measure effects on patients who do not receive surgery as on patients who do. In short, COGNATE assessed whether the addition of EUS to usual staging tests changes treatment, improves survival and quality of life, and uses resources cost-effectively. The COGNATE team also developed a quality assurance process for EUS scans.

Methods

Design and interventions
COGNATE was a pragmatic, eight-centred, two-arm randomised controlled trial. All patients with gastro-oesophageal cancer received standard staging algorithms, after which the relevant multidisciplinary team (MDT) chose a provisional management plan from: endoscopic mucosal resection (EMR); immediate surgery; surgery after neo-adjuvant chemotherapy; and chemotherapy and radiotherapy. In principle patients randomised to the intervention group then received EUS, while those randomised to the control group continued with their agreed management plan.

Setting

The trial took place in eight British hospitals, two of which – Aberdeen Royal Infirmary, a teaching hospital, and Gloucestershire Royal Hospital, a district general hospital – contributed most participants.

Participants

Participants were eligible for the trial if they had cancer of the oesophagus, stomach or gastro-oesophageal junction and had not started treatment. To be randomised, patients had to be free of metastatic disease, fit for surgery (even if not planned) and graded less than 3 on both ASA (American Society of Anesthesiologists) and WHO (World Health Organization) criteria. Clinicians could exclude from the trial any patient for whom they were not in equipoise about the value of EUS. The trial co-ordinating centre monitored all exclusions and reasons given.

We invited eligible patients to participate in the trial, gave them the patient information sheet approved by the Multi-centre Research Ethics Committee for Scotland, and allowed them time to study it and ask questions. We stressed that the choice of treatment after EUS was the same in both groups. We then asked consenting patients to sign the approved consent form.

Randomisation

We stratified consented patients by centre and tumour location, and randomised them in equal proportions between EUS and not – by telephone call to the co-ordinating centre, which used dynamic software to prevent subversion.

Sample size

Our original application proposed survival as the primary outcome and a target of 700 patients. Difficulty in recruiting centres led us to change the primary outcome to 'quality-adjusted survival' and the target to a maximum of 400 and a minimum of 220. The latter would yield 80% power to detect a hazard ratio of 0.5, or a 'small' effect size of 0.4 in quality of life, or some combination of these.

Follow up

We followed participants until death or the end of data collection, which was between 12 and 54 months after recruitment. We collected data at discharge from hospital after initial treatment and at follow-up clinics after 1, 3, 6, 12, 18, 24 and 36 months.

Outcome measures

To compare the two randomised groups we used:

(a) quality-adjusted survival (primary outcome)
(b) survival censored at between 12 months (for those last recruited) and 54 months
(c) participant-reported quality of life using three questionnaires: European Quality of Life – 5 Dimensions (EQ-5D) (generic), Functional Assessment of Cancer Therapy – General (FACT-G) scale (cancer related) and FACT Additional Concerns (FACT-AC) scale (gastro-oesophageal cancer specific)
(d) process of care:
   ○ changes in management plans agreed by MDTs
   ○ complete resection rate, and
   ○ adverse events related to EUS
(e) use of health-care resources.

Psychometric methods to refine quality-of-life measurement

We asked participants to assess their quality of life through EQ-5D and FACT, specifically the general module (FACT-G), the oesophageal module [Functional Assessment of Cancer Therapy – Oesophageal (FACT-E)] and the gastric module [Functional Assessment of Cancer Therapy – Gastric (FACT-Ga)]. As FACT-E and FACT-Ga have many similar questions, the FACT team encouraged us to combine them into a single 'Additional Concerns' module. We used factor analysis to examine the structure of FACT and thereby

assess whether to aggregate these two modules into one. We used structural equation modelling to examine the relationships between EQ-5D and FACT scores, and thereby assess whether EQ-5D reflected all aspects of quality of life experienced by these patients.

Statistical methods

Primary analysis was by allocated group, whether or not participants received EUS. This reflects the pragmatic nature of the trial and our aim of evaluating EUS in informing decisions in the real world. We used baseline characteristics, including quality-of life scores, as covariates to improve the precision and generalisability of the model. Although blinding participants to their allocation was neither possible nor desirable, those responsible for analysis remained blind until the Trial Steering Committee and Data Monitoring and Ethics Committee had reviewed the definitive analysis.

Economic methods

We assessed the cost-effectiveness of EUS in the diagnosis and treatment of gastro-oesophageal cancer by estimating differences between the cost of patients' care including EUS, and the cost when limited to conventional staging; and corresponding differences in effectiveness as estimated by quality-adjusted life-years (QALYs). We used 'bootstrapping' to overcome the skewed data and cost-effectiveness acceptability curves to quantify uncertainty.

In accordance with National Institute for Health and Care Excellence (NICE) guidance we analysed COGNATE from the perspective of the NHS, focusing on health-care resources used by participants after randomisation. These included investigation, treatment and palliation, and other elements of secondary and pharmaceutical care. The local co-ordinator at each trial site used an electronic database to record the use of NHS resources by participants throughout the period of the trial. To cost these resources, we refined and used published national unit costs.

We used sensitivity analysis to explore whether or not the estimated benefits and costs of endoscopic ultrasonic staging relative to conventional staging were sensitive to key features of our analysis, notably the cost of EUS scans.

Quality assurance of endoscopic ultrasound scans

The COGNATE team asked trial sites to record EUS scans as videos and their interpretations on a proforma. A panel of six COGNATE investigators reviewed 20 anonymised scans (21% of 97 performed within COGNATE trial), both individually before, and together during, five web-based conferences. Each conference reached a blinded consensus on the staging of four tumours. We compared the original report with the staging decisions of individual reviewers, their consensus and that of an external reviewer.

Results

We randomised 223 patients, of whom 213 (96%) yielded enough data for primary analysis. Over three-quarters of participants were male, and nearly half were over 65 years. The most common tumour site was the oesophagus and the most common tumour type was adenocarcinoma. The most common tumour stage was T3, and slightly more participants had nodal stage N1 than N0. At the end of the trial 44% of EUS participants and 32% of control participants were alive.

Our psychometric analyses confirmed that it was appropriate to aggregate the FACT-E and FACT-Ga modules into a single Additional Concerns module. The structural equation modelling suggested that EQ-5D scores captured individuals' physical and functional well-being but not their social and emotional well-being as measured by FACT. There are two possible explanations for this: either EQ-5D items do not cover these domains; or the general public in weighting the health states defined by EQ-5D chose to give little weight to social and emotional issues. However EQ-5D utility scores correlated well with EQ-5D visual analogue scores. This is consistent with patients and the public agreeing that physical and functional

well-being deserve more weight, possibly because of the extreme nature of gastro-oesophageal cancer. Thus our preparatory analysis supported our use of EQ-5D as a generic measure of quality of life, and persuaded us to adjust survival using not only EQ-5D, but also FACT-G and FACT-AC.

Endoscopic ultrasound significantly improved participant survival, with a hazard ratio of 0.706 [95% confidence interval (CI) from 0.501 to 0.996] and an increase of 121 days in estimated median survival – from 1.63 years in the control group to 1.96 years in the intervention group. Participants reported consistent, although non-significant, improvements in mean outcomes at 12 months, notably a difference of 0.061 (95% CI from −0.043 to 0.164) in mean EQ-5D scores between 0.449 in the control group and 0.509 in the intervention group; and a difference of 0.12 (95% CI from −0.27 to 0.51) in mean FACT-G between 2.15 in the control group and 2.27 in the intervention group. Combining survival and quality of life, EUS improved survival adjusted for generic quality of life with a hazard ratio of 0.705 (95% CI from 0.499 to 0.995) and an increase of 66 days in estimated median quality-adjusted survival – from 0.94 QALYs in the control group to 1.12 QALYs in the intervention group.

Trial sites reported consistent, although non-significant, reductions in total resource use in secondary and pharmaceutical care (including EUS scans when undertaken), generating mean savings of about £2860 (95% 'bootstrapped' CI from −£2200 to £8000) from an average of £32,000 [with a standard deviation (SD) of £22,000] in the control group to £29,200 (SD £14,900) in the intervention group. Combining these estimated benefits and savings yields probability of 96.6% that EUS is cost-effective in the sense of achieving the NICE criterion of costing less than £20,000 to gain a QALY.

The benefits of EUS were significantly greater for those with poor initial quality of life, but did not differ between centres. Similarly there was a significant interaction between initial quality of life and the effect of EUS on all the FACT scales; again, sicker patients benefitted more from EUS. However there was no significant difference between intervention and control groups in mean FACT scores adjusted for covariates. There were no serious adverse reactions attributable to EUS.

Both management plans and final treatment varied between centres. EUS increased the proportion of tumours completely resected from 80% (44 out of 55) to 91% (48 out of 53). Furthermore participants allocated to EUS who then transferred to a 'therapeutic' treatment, namely EMR or surgery in some form, survived much better than control subjects who made this change, and better than intervention participants confirmed for one of these 'therapeutic' treatments. The few intervention participants who transferred to 'conservative' treatment, namely chemotherapy, radiotherapy or both, survived worse than both the control subjects who made this change and intervention participants confirmed for 'conservative' treatment. In contrast, control subjects who changed plans in either direction experienced intermediate survival, arguably because they lacked the discriminatory power of EUS. In short, changes are consistently more appropriate in the intervention group than in the control group. Although all of these analyses have low power, and are therefore not statistically significant, they underpin the significant differences in outcomes, and help to explain them.

The quality assurance panel achieved excellent agreement on the tumour node metastasis (TNM) staging of the 20 endoscopic ultrasonic films. Their web-based conferences failed to agree staging for only one T stage and only one N stage, and reached consensus agreeing with the original investigator on 19 T stages and 17 N stages. There was excellent agreement between the original investigator and the consensus on T stage (weighted kappa = 0.866; $p < 0.001$) and moderate agreement on N stage (kappa = 0.562; $p = 0.012$). In short, we developed an effective quality assurance process for EUS scans. It provides a useful model for future NIHR-funded assessments of diagnostic technologies and has the potential to improve routine clinical practice.

Conclusions

Although EUS was common practice by the time COGNATE began recruiting in 2005 we achieved the first rigorous evaluation of EUS. The COGNATE team ameliorated many problems of recruiting centres and participants, mainly through two centres – Aberdeen and Gloucester. EUS achieved a surprising combination of significant improvements in survival (121 days) and quality-adjusted survival (66 days); a substantial, although non-significant, net saving of £2800 per trial participant; and, combining these statistical and economic findings, 96.6% probability of being cost-effective by NICE criteria. We judge that these impressive findings provide strong evidence in favour of EUS scans for all patients with gastro-oesophageal cancer who have the potential to benefit.

As the COGNATE team 'caught the EUS horse just as it was leaving the stable', we make no recommendation for future research into EUS. Instead we recommend:

1. policy-orientated research into the best time to evaluate new technologies
2. methodological research to streamline the collection of data to evaluate complex technologies such as EUS, notably on the costs of the extensive care for conditions such as gastro-oesophageal cancer, and
3. psychometric research to refine the integrated Functional Assessment of Cancer Therapy – Oesophageal and Gastric (FACT-EG) module as a valid measure of the outcome of gastro-oesophageal cancer.

Trial registration

This trial is registered as ISRCTN1444215.

Funding

Chapter 1 Introduction

Background

In the UK, cancer of the upper gastrointestinal tract (oesophageal or gastric, or both) affects some 13,000 patients each year. Gastro-oesophageal cancer is the fifth most frequently occurring cancer in the UK and the fourth most common cause of cancer death.[1,2]

Many Western series have described recent changes in the epidemiology of gastro-oesophageal cancer, characterised by reduced incidence of distal gastric cancer and increased incidence of proximal gastric and distal oesophageal cancer.[3] Furthermore the incidence of these cancers varies between regions, with more oesophageal adenocarcinoma in Scotland (18 per 100,000) than in England (13 per 100,000).[1,4]

Most community-based series show that cancer of the stomach or oesophagus mostly affects the elderly and often causes significant morbidity. The Scottish Audit of Gastric and Oesophageal Cancer (SAGOC)[4] reported a median age of 72 years for patients with gastric or oesophageal cancer; this was similar to that in the recent National Oesophago-Gastric Cancer Audit (NOGCA) in England and Wales, initiated by the Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland (AUGIS).[5] In both studies, tumours were unusual in patients who were <40 years old. About 40% of patients with upper gastrointestinal cancer have significant comorbid disease at presentation and about one-sixth are in bed for more than half of the time.

The general prognosis of patients with gastro-oesophageal cancer is poor, with a median survival after diagnosis among all patients in 1997–99 of 8 months.[4] Although survival had improved by 2005, 5-year survival remained poor at 7% for oesophageal cancer and 12% for gastric cancer.[6] However survival depends on tumour stage and patient characteristics, and there have been many advances in the treatment of these tumours, both curative and palliative. So it is important to select the most appropriate management plan for each patient.

Endoscopic ultrasound (EUS; or endosonography) is a medical procedure performed by gastroenterologists, radiologists or surgeons with specialised training. Endosonography combines endoscopy – the insertion of a probe into the upper gastrointestinal tract – with ultrasonography. It places a high-frequency ultrasound probe mounted on the end of the endoscope in direct contact with oesophageal or gastric tumours. This provides good images of the structures of the bowel wall and local lymph nodes, but is less good at identifying distant metastases. To patients it feels very similar to normal endoscopy, unless it includes ultrasound-guided biopsy of deeper structures. Although biopsy may increase risk, the basic procedure is no more risky than an endoscopy.

The literature review (see *Literature review*) shows that EUS has potential to provide accurate staging of gastric and oesophageal tumours, rather than associated nodes. It can therefore provide important prognostic information to guide management. However, as the link between better staging and better management is not proven, the benefit of EUS is not clear.

So we designed the trial known as Cancer of Oesophagus or Gastricus: New Assessment of Technology of Endosonography (COGNATE) to evaluate, not the accuracy of EUS, but the effect it has on patient management and thus outcome. Accordingly the choice of treatment was an important intermediate outcome. It was also crucial to follow patients with gastro-oesophageal cancer who had not been selected for surgery as well as those selected; if EUS leads to more or less surgery, it is as important to measure effects on patients who do not receive surgery as on patients who do. Although this comprehensive

approach is a central feature of COGNATE, the literature review (see *Literature review*) shows that many assessments of the effect of staging lack this breadth.

The COGNATE trial therefore monitored the outcome of treatment to detect increased mortality or morbidity. If it generates evidence that EUS improves choice of treatment, this will benefit, not only individual patients, but the whole population of patients with gastro-oesophageal cancer, through better targeting of scarce resources. Thus the National Coordinating Centre for Health Technology Assessment (NCCHTA) commissioned COGNATE to evaluate whether EUS is effective and cost-effective in the management of gastro-oesophageal cancer. There is no other published or current randomised trial that addresses this issue of importance to the care of cancer in the NHS. In short, COGNATE aims to estimate the value of EUS in managing gastro-oesophageal cancer.

Literature review

Treatment

Endoscopic treatment

For patients with early gastro-oesophageal cancer, endoscopic treatments, notably endoscopic mucosal resection (EMR), can achieve long-term cure without the risk and morbidity of surgery. The risk of a major complication (e.g. perforation or bleeding) is approximately 1%.[7] If the tumour is localised to the mucosa, EMR is likely to lead to long-term survival with a 5-year disease-free survival rate of 99% and a general 5-year survival rate of 84%.[8] Even in patients with early submucosal changes, EMR may be the treatment of choice. As the tumour invades deeper, however, there is an increased risk of lymphatic involvement needing a surgical approach.[9–11]

Surgery

Surgery for gastro-oesophageal cancer is a major therapeutic intervention with substantial postoperative morbidity and mortality. Even in patients surviving surgery, quality of life deteriorates and may take several months to recover to the preoperative state. Indeed patients who die within 2 years of oesophageal surgery seldom recover their preoperative quality of life.[12,13] Hence it is important to restrict surgery with curative intent to patients likely to achieve long-term survival. Both the ability of surgeons to achieve complete resection of the tumour (R0) by removing all macroscopic and microscopic lesions and the outcome of that surgery depend on the fitness of the patient and the extent of the tumour at the time of surgery.

Patients in whom complete resection is possible have a significant survival advantage over those whose resections are incomplete.[4] Indeed incomplete resection of oesophageal cancers increases neither length nor quality of survival.[14] The most common reason for incomplete resection in patients with oesophageal cancer is residual tumour in the resection margins.[15,16] The presence of metastases in lymph nodes also reduces general and disease-free survival.[4]

Neo-adjuvant therapy

The development of effective chemotherapy regimens for patients with advanced disease has led to the introduction of neo-adjuvant chemotherapy before surgery for both oesophageal and gastric cancer. A Cochrane review of neo-adjuvant chemotherapy in oesophageal cancer, based on 2000 patients in 11 randomised controlled trials (RCTs), showed a survival advantage at 5 years for chemotherapy before surgery compared with surgery alone.[17] However the two largest trials included in this review yielded conflicting results. The Medical Research Council (MRC)-funded trial of fluorouracil (5FU) and cisplatin before surgery compared with surgery alone showed a median survival advantage of 4 months in the neo-adjuvant arm.[18] However a similar trial from the USA[19] failed to show any effect of neo-adjuvant chemotherapy. Nevertheless meta-analysis of all trials shows a survival advantage after 5 years for neo-adjuvant chemotherapy in operable oesophageal cancer.

In gastric cancer, meta-analysis of trials of chemotherapy after surgery compared with surgery alone showed a small survival benefit from chemotherapy, particularly for lymph-node-positive disease; however toxicity was significant.[20] A large RCT comparing surgery alone with chemoradiotherapy after surgery showed a survival benefit from chemoradiotherapy, but sustained criticism for not controlling the quality of surgery.[21] The MRC-funded trial of chemotherapy before and after surgery compared with surgery alone showed significantly better 5-year survival in the chemotherapy group – 36% compared with 23%.[22]

Multimodal treatment

Although surgery-based treatment remains the norm for the treatment of gastro-oesophageal cancer, external-beam radiotherapy alone has achieved excellent results for oesophageal cancer.[23–25] Concomitant chemoradiotherapy is also effective for oesophageal cancer.[26–28] However there is no evidence that external-beam radiotherapy alone is adequate for gastric cancer. The few RCTs comparing external-beam radiotherapy with surgery alone have been underpowered. Similarly, the two RCTs that have compared chemoradiotherapy before surgery with chemoradiotherapy alone have not shown any benefit to either treatment.[29,30] A study from China[31] showed no difference between surgery and chemoradiotherapy alone for the treatment of squamous cell oesophageal cancer.

In advanced localised oesophageal cancer there is good evidence that chemoradiotherapy is superior to external-beam radiotherapy alone in achieving long-term survival and improving swallowing.[28,32,33] In gastric cancer, however, radiotherapy is more difficult.

Staging and treatment selection

It is clear from both SAGOC[4] and NOGCA[5] that there is variation in the selection of patients for different treatments. Operation rates in the Scottish audit varied by tumour type: oesophageal cancer 31%; gastro-oesophageal junction tumours 38%; and gastric cancers 51%. There was even greater variation between centres in operation rates: oesophageal cancer 20–42%; junction cancer 25–64%; and gastric cancer 32–63%.[4] The NOGCA report[5] showed similar variation in patients selected for curative surgery.

The general criteria for treatment selection are stage of the tumour at presentation, along with patient's fitness, and ability and willingness to undergo specific treatments. Patient fitness depends on comorbid disease. Management decisions depend on the interaction of all of these factors. In addition they increasingly depend on markers of the biological behaviour of a tumour. For gastro-oesophageal cancer these may include tumour differentiation, growth characteristics, response to chemotherapy and molecular mechanisms.

Tumour staging

Accurate assessment of tumour stage at presentation will inform subsequent management decisions by indicating likely prognosis and the feasibility of specific treatments.

For gastro-oesophageal cancer the issues are:

- Is EMR treatment likely to be possible?
- Does surgical resection have a high probability of complete resection?

Staging is also important for meaningful comparisons between trials. Much of the uncertainty in comparing treatment options arises from inaccurate staging.

Tumour staging summarises anatomical measurements of the extent of direct invasion by a tumour (T); the involvement of lymph nodes (N); and the presence of distant metastases (M). The most common staging investigations used for patients with upper gastrointestinal cancer are computed tomography (CT) scanning, and EUS. In SAGOC[4] 69% of patients received CT. In NOGCA[5] most patients received CT but the rate fell with increasing age and in patients with poor performance status. Magnetic resonance imaging

(MRI) and integrated positron emission tomography and computed tomography (PET-CT) scanning are less common. For gastric cancer, laparoscopy can add considerably to these imaging techniques.

Endoscopic ultrasound
Endoscopic ultrasonography was introduced in the early 1980s. However it became accepted practice only in the 2000s. For example the EUS rate was 3% in SAGOC undertaken in the late 1990s,[4] but 48% (gastric) and 58% (oesophageal) in NOGCA[5] undertaken some 10 years later. In this review we consider the accuracy of EUS for both gastric and oesophageal cancers, and for both T and N stages.

Search strategy
The first systematic review of EUS in gastro-oesophageal cancer dates from 2001.[34] Two recent updates on EUS in oesophageal[35] and gastric cancer[36] reviewed studies up to 2006. As these reviews used similar search strategies, we used that strategy to identify articles up to October 2009. Specifically, we searched MEDLINE, PubMed, Ovid journals and The Cochrane Library for articles including all the following terms: endoscopic ultrasound *or* endosonography; oesophageal cancer *or* gastric cancer; tumour staging; invasion *and* surgery. We excluded studies that did not define tumour location clearly, those that did not confirm findings by surgery and those with fewer than 10 patients. Of studies from the same centre reporting the same data, we included only the most up-to-date reports. In contrast to previous systematic reviews, we classified tumours of the gastro-oesophageal junction as oesophageal cancers.

Gastric cancers
We identified 29 studies, with 2500 patients, that reported on the accuracy of EUS staging for gastric cancer between 1988 and 2009;[37–65] 21[37–41,43–45,47,48,50–53,56,57,59,62–65] used radial ultrasound probes, three[49,58,60] used linear array probes and five[42,46,54,55,61] did not report the type of probe (*Table 1*). As the reported accuracy of linear array probes for both T and N stages did not differ from that of the rest, we include them here.

The pooled accuracy of EUS for gastric T stage was 76.2% (range 35–92%). Accuracy was 71.2% for studies reported before 2000, and slightly but not significant less than 80.4% for studies after 2000. This is consistent with Puli *et al.*,[36] who reported no difference between studies published in the 1980s, 1990s or 2000s.[36] Their review also compared the sensitivity and specificity of EUS at different T stages of gastric cancer with the pathology from resected specimens in 22 studies with 1900 patients (*Table 2*).

Thus sensitivity for tumour invasion is high for T1, lower for T2, and then improves as tumours become more advanced. In contrast, specificity is very high for all stages of disease, but highest for T1. Hence if EUS shows T1 disease, the patient probably has anatomical T1 disease. In contrast, if EUS shows T2 disease, the patient may have anatomical T1 disease. So EUS can result in overtreatment, subjecting patients to resectional surgery rather than EMR in the first instance.

The pooled diagnostic accuracy of EUS for nodal staging of gastric cancers was 67.9% (range 42–90%), lower than for T stage as reported in previous studies.[34,36] However it is likely that the use of linear array probes and fine-needle cytology increases that accuracy.

Comparison of endoscopic ultrasound and computed tomography
Six studies compared the diagnostic accuracy of CT with that of EUS for both T and N stages (*Table 3*). Diagnostic accuracy for both T and N staging was higher for EUS than for CT. However the last two studies showed high levels of diagnostic accuracy for both EUS and CT, probably because they were the only studies to use multi-detector row-computed tomography (MDCT). The superiority of MDCT over conventional CT is also apparent in a review by Kwee and Kwee,[66] assessing different imaging modalities for lymph node status in gastric cancer. Nevertheless this review concluded that neither EUS nor MDCT reliably excluded or confirmed the presence of lymph node metastases in gastric cancer. In a separate review, Kwee and Kwee[67] concluded that EUS was the best imaging modality for T staging of gastric cancer.

TABLE 1 Studies reporting accuracy of EUS in staging gastric cancers

Study	Design	Number of patients	Transducer	Accuracy (%) T stage	Accuracy (%) N stage
Murata 1988[53]	Prospective	146	Radial	79	88
Tio 1989[59]	Prospective	72	Radial	84	66
Botet 1991[43]	Prospective	50	Radial	92	78
Saito 1991[57]	Prospective	110	Radial	86	Unspecified
Akahoshi 1991[38]	Prospective	74	Radial	81	50
Rosch 1992[65]	Consecutive	41	Radial	71	75
Grimm 1993[48]	Prospective	147	Radial	78	Unspecified
Dittler 1993[45]	Consecutive	254	Radial	83	66
Ziegler 1993[64]	Prospective	108	Radial	86	74
Caletti 1993[44]	Prospective	42	Radial	91	69
Perng 1996[54]	Consecutive	69	Unspecified	71	65
Massari 1996[52]	Prospective	65	Radial	89	68
Francois 1996[46]	Consecutive	29	Unspecified	79	79
Hunerbein 1996[49]	Consecutive	60	Linear	65	73
Wang 1998[61]	Consecutive	119	Unspecified	70	65
Willis 2000[62]	Consecutive	116	Radial	78	77
Xi 2003[63]	Prospective	35	Radial	80	69
Shimoyama 2004[58]	Consecutive	45	Linear	71	80
Polkowski 2004[55]	Prospective	88	Unspecified	63	47
Javaid 2004[50]	Consecutive	112	Radial	83	64
Bhandari 2004[42]	Prospective	48	Unspecified	88	79
Tsendsuren 2006[60]	Consecutive	41	Linear	68	66
Arocena 2006[40]	Prospective	17	Radial	35	42
Potrc 2006[56]	Prospective	82	Radial	68	57
Ang 2006[39]	Prospective	57	Radial	77	60
Ganpathi 2006[47]	Consecutive	109	Radial	80	78
Bentrem 2007[41]	Prospective	225	Radial	57	50
Lok 2008[51]	Prospective	123	Radial	64	75
Ahn 2009[37]	Prospective	68	Radial	90	90

Influence of endoscopic ultrasound on management

Although the accuracy of EUS in staging gastric cancer is thus well reported, there are few studies examining the effect of EUS staging on subsequent management. Dittler and Siewert[45] found that EUS predicted complete (R0) resection of gastric tumours in 81% of 254 consecutive gastrectomies, close to the achieved complete resection rate of 78%. Javaid *et al.*[50] also describe a high complete resection rate in patients predicted by EUS. A Chinese study of only 35 patients reported that the sensitivity of EUS for resection rates was 88% and the specificity 100%.[63]

TABLE 2 Pooled sensitivity and specificity (%) by T stage for patients with gastric cancer. Reproduced with permission from Puli SR, Reddy JBK, Bechtold ML, Antillon MR, Ibdah JA. How good is endoscopic ultrasound for TNM staging of gastric cancers? A meta-analysis and systematic review. *World J Gastroenterol* 2008;**14**:4011–19.

T stage	Pooled sensitivity (range)	Pooled specificity (range)
T1	88.1 (84.5–91.1)	100 (99.7–100)
T2	82.3 (78.2–86.0)	95.6 (94.4–96.6)
T3	89.7 (87.1–92.0)	94.7 (93.3–95.9)
T4	99.2 (97.1–99.9)	96.7 (95.7–97.6)

Source: Puli *et al.* 2008[36]

TABLE 3 Accuracy of T and N staging in patients with gastric cancer: EUS vs CT

Study	Number of patients	Stage	EUS (%)	CT (%)
Botet 1991[43]	50	T	92	42
		N	78	48
Ziegler 1994[64]	108	T	86	43
		N	74	51
Perng 1996[54]	69	T	71	42
		N	65	49
Polkowski 2004[55]	88	T	63	44
		N	47	52
Bhandari 2004[42]	48	T	88	83
		N	79	75
Ahn 2009[37]	68	T	90	86
		N	90	84

However the accuracy of EUS for T1 tumours is less than for T4 tumours. With conventional 7.5-Hz or 12-Hz endoscopic transducers it is difficult to determine whether a tumour is limited to the gastric mucosa or invading the submucosa and to what extent. It is this distinction that enables EUS to predict the success of EMR. The review by Kwee and Kwee[68] of the few studies that address this point is uncertain whether EUS can accurately differentiate between mucosal and deeper gastric cancers.

Oesophageal cancers

We identified 40 studies, with 2600 patients, which reported on the accuracy of EUS staging for oesophageal cancer between 1986 and 2009 (*Table 4*).[48,49,53,59,65,69–103] The pooled accuracy for T stage was 78.5% (range 59–93%) and for N stage 76.3% (range 60–90%). As with gastric cancer, there was no significant evidence that accuracy improved with time. Indeed the studies with the highest diagnostic accuracy were all undertaken before 2000.

In a systematic review of EUS in the staging of gastro-oesophageal cancer, Kelly *et al.*[34] found that non-traversability of oesophageal cancers and tumours at the gastro-oesophageal junction reduced diagnostic accuracy of EUS staging, but not significantly. In contrast Hordijk *et al.*[74] found that accuracy was about 90% whether tumours were traversable or not, but fell to 46% for tumours that had been dilated. Accordingly we postulate that differences between studies may arise from the percentage of non-traversable tumours and the method of dealing with these. Two other studies[78,85] found a significant

TABLE 4 Studies reporting on the accuracy of EUS in staging oesophageal cancers

Study	Design	Number of patients	Accuracy (%)	
			T stage	N stage
Murata 1988[53]	Consecutive	173	88	88
Tio 1989[59]	Prospective	91	90	82
Vilgrain 1990[69]	Consecutive	46	73	Unspecified
Botet 1991[70]	Prospective	50	92	88
Rice 1991[71]	Consecutive	22	59	70
Ziegler 1991[72]	Prospective	37	89	69
Rosch 1992[65]	Consecutive	44	82	70
Dittler 1993[73]	Consecutive	167	86	73
Grimm 1993[48]	Prospective	63	86	86
Hordijk 1993[74]	Consecutive	41	76	Unspecified
Yoshikane 1994[77]	Consecutive	28	75	72
Greenberg 1994[75]	Prospective	16	85	60
Peters 1994[76]	Consecutive	34	76	82
Binmoeller 1995[78]	Prospective	38	89	79
McLoughlin 1995[79]	Consecutive	15	86	Unspecified
Hasegawa 1996[80]	Consecutive	22	76	67
Holden 1996[81]	Consecutive	15	87	73
Hunerbein 1996[49]	Consecutive	19	84	88
Massari 1997[83]	Prospective	55	90	87
Natsugoe 1996[82]	Consecutive	37	Unspecified	86
Pham 1998[84]	Consecutive	28	61	75
Vickers 1998[85]	Prospective	50	92	86
Bowrey 1999[86]	Prospective	30	93	80
Catalano 1999[87]	Prospective	145	Unspecified	73
Nishimaki 1999[88]	Consecutive	166	Unspecified	72
Salminen 1999[89]	Consecutive	26	66	72
Heidemann 2000[90]	Consecutive	68	79	79
Nesje 2000[91]	Prospective	54	70	90
Vazquez 2001[92]	Consecutive	64	Unspecified	70
Kienle 2002[93]	Prospective	117	69	79
Chang 2003[94]	Prospective	60	83	89
Wu 2003[95]	Prospective	31	84	71
DeWitt et 2005[96]	Prospective	102	72	75
Lowe 2005[97]	Prospective	75	71	81
Moorjani 2007[99]	Prospective	50	64	72
Shimpi 2007[100]	Prospective	42	76	89
Kutup 2007[98]	Prospective	214	66	64
Sandha 2008[102]	Prospective	16	80	81
Mennigen 2008[101]	Prospective	97	73	74
Takizawa 2009[103]	Prospective	159	Unspecified	64

reduction in diagnostic accuracy in stenosed oesophageal tumours and suggested that this may be because all the tumours were T3 or T4 with a high rate of lymph node involvement. Kelly et al.[34] also identified junctional tumours as a potential, but not statistically significant, source of diagnostic inaccuracy owing to the difficulty in getting contact between the probe and the tumour surface. As few studies report the exact location of tumour sites, there is uncertainty whether this is a confounding variable.

As with gastric cancer, the accuracy of EUS was better in more advanced oesophageal cancers; specificity is high at all tumour stages, whereas sensitivity increases for T3 and T4 tumours (*Table 5*).[35] This suggests a slight tendency for EUS also to overestimate oesophageal T stage.

Two meta-analyses[35,104] have reported sensitivities above 80% and specificities about 80% for EUS in estimating lymph node involvement in patients with oesophageal cancer. Puli et al.[35] also found four studies that combined EUS with fine-needle aspiration cytology (FNAC) and increased sensitivity to 97% (range 92–99%) and specificity to 95% (range 91–98%). In contrast, Van Vliet et al.[104] found no improvement in accuracy by combining FNAC and EUS. However they did find five studies that analysed results for mediastinal lymph nodes separately from those for coeliac lymph nodes, which had a higher sensitivity of 85% (range 72–99%) and a higher specificity of 96% (range 92–100%). It is likely that these high accuracies for FNAC and coeliac nodes were in specialised centres. The identification of coeliac lymph node involvement may have greater potential to improve management as this suggests metastatic disease in patients with oesophageal cancer, and thus precludes surgery.

Endoscopic ultrasound compared with computed tomography scanning

Endoscopic ultrasound staging consistently has higher diagnostic accuracy than CT staging for both T and N stages (*Table 6*). Unlike gastric cancer, there is no reported comparison of EUS with the most up-to-date CT techniques. Although there are few studies comparing EUS with MRI or PET scanning, these are not generally more accurate than CT for local regional upper gastrointestinal cancers.

Influence of endoscopic ultrasound on management

The accuracy of EUS staging is as well reported for oesophageal cancer as for gastric cancer. Again, however, there are few studies examining the effect of EUS staging on subsequent management. In their systematic review, Dyer et al.[105] acknowledged that drawing conclusions from observational studies, rather than RCTs, was open to bias, but estimated that EUS appeared to change management in 24–29% of patients. Two retrospective,[106,107] and therefore suspect, studies examined the effect of EUS staging on patient survival. The first study[106] reported significantly better survival and reduced recurrence rate following better selection of patients for surgery and neo-adjuvant treatment; although the second study[107] found no advantage from EUS staging, it omitted to report on patients declined for surgery. Neither study reported on quality of life.

TABLE 5 Sensitivity and specificity (%) by T stage for patients with oesophageal cancer. Reproduced with permission from 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**:1479–90.

T stage	Pooled sensitivity (range)	Pooled specificity (range)
T1	81.6 (77.8–84.9)	99.4 (99.0–99.7)
T2	81.4 (77.5–84.8)	96.3 (95.4–97.1)
T3	91.4 (89.5–93.0)	94.4 (93.1–95.5)
T4	92.4 (89.2–95.0)	97.4 (96.6–98.0)

Source: Puli et al. 2008.[35]

TABLE 6 Accuracy of staging in oesophageal cancer: EUS vs CT vs MRI vs PET

Study	Number of patients	Stage	EUS (%)	CT (%)	MRI (%)	PET (%)
Botet 1991[70]	50	T	92	60		
		N	88	74		
Ziegler 1991[72]	37	T	89	51		
		N	69	51		
Greenberg 1994[75]	28	T	85	15		
		N	60	50		
Holden 1996[81]	15	T	87	40		
		N	73	33		
Massari 1997[83]	55	T	90	50		
		N	87	39		
Kienle 2002[93]	117	T	69	33		
		N	79	67		
Wu 2003[95]	31	T	84	68	60	
		N	71	78	64	
Lowe 2005[97]	69	T	71	42		42
		N	81	80		76
Sandha 2008[102]	16	T	80			
		N	81	69		56

Summary

Many studies have assessed the accuracy of EUS in the staging of gastro-oesophageal cancer. The pooled rates for T stage suggest accuracy of 76% for gastric cancer and 78% for oesophageal cancer; those for N stage suggest accuracy of 68% for gastric cancer and 76% for oesophageal cancer. Furthermore accuracy improves for more advanced tumours. These estimates of accuracy are consistently better than those achieved with other imaging modalities, most often CT scanning. However there is little rigorous evidence as to whether increased accuracy translates into improvements in patient management, still less patient outcomes.

However the management of gastro-oesophageal cancer depends, not only on staging accuracy, but also on patient factors like fitness and willingness for treatment, and treatment factors like benefits and risks. Many patients with gastro-oesophageal cancer have substantial comorbid disease. Often this determines treatment selection irrespective of tumour stage. For other patients the differentiation between T2 and T3 tumours may have little influence on treatment selection, as it is unclear how this affects prognosis or whether treatment should differ between these tumours. So increased accuracy from EUS may be most valuable in discriminating between T3 and T4 tumours and judging whether complete resection is feasible; and between T1 and T2 tumours and judging whether endoluminal treatment is feasible.

Nodal status is another important prognostic indicator for gastro-oesophageal cancer, but it is less clear how this should affect management decisions. Patients with tumours that are N-positive and T3 or T4 will generally fare worse than those with less advanced tumours. Nevertheless we do not know whether and how the outcome for patients with more advanced tumours depends on the choice between curative and palliative treatment.

In short, the known accuracy of EUS in staging gastro-oesophageal cancer makes it important to evaluate whether this staging modality significantly affects the management of gastro-oesophageal cancer. Only a RCT of all patients eligible for EUS can answer that question.

Philosophy

Evaluative paradigm

There has been no formal evaluation of EUS, merely recommendations that it was essential in staging oesophageal cancers.[108] Nevertheless the 2008 NOGCA[5] showed that even cancer networks do not universally use EUS to stage oesophageal cancers. Staging non-traversable tumours is difficult;[34] the majority are T3 or T4 lesions, which need good staging to avoid non-curative resections. EUS is least accurate in carcinomas around the gastro-oesophageal junction,[34] incidence of which is increasing rapidly. Another problem is that there are few studies comparing the cost-effectiveness of EUS with that of modern CT protocols.[34]

In summary, although there is evidence that EUS improves anatomical staging of patients with gastro-oesophageal cancer, it is not clear how it affects patient management, even less how it affects patient outcome. SAGOC showed that between 1997 and 2000 few patients with gastro-oesophageal cancer underwent EUS.[4] The subsequent growth in use of EUS in gastro-oesophageal cancer, documented in NOGCA, reinforces the case for evaluating the contribution of EUS to management.

It is also important to study which patients with gastro-oesophageal cancer benefit from EUS. At first sight three types of cancer have the best chance:

1. T1 tumours localised to the mucosa, in which endoscopic treatment may avoid unnecessary surgery.
2. Tumours for which EUS may predict the outcome of 'curative' surgery, in particular the risk of residual disease.
3. T3 or T4 tumours in which EUS may encourage multimodal treatment, taking the form of chemotherapy alone, radiotherapy alone, both or neither, depending on clinical circumstances.

To address all these issues comprehensively needs a pragmatic randomised trial that assesses patients by a conventional staging algorithm and then randomises them between EUS and not. In designing such a trial, we started from the seminal writing of Schwartz and Lellouch[109] (*Table 7*).

At a time when randomised trials were much rarer than today, Schwartz and Lellouch prepared the way for 'health technology assessment' by distinguishing between two distinct *scientific* paradigms for clinical trials: 'fastidious' trials aim to test defined scientific hypotheses and 'pragmatic' trials aim to choose between alternative *technologies.*[109]

In practice, trials that keep to either column of *Table 7* are rare. For example the proposal that pragmatic trials need no significance test is feasible only if the protocol specifies how analysis will combine the potential criteria to yield a single decision function, and there is enough information about that function to ensure that the resulting sample size calculation yields the required statistical confidence in the simple decision to choose the technology that performed best on that function in the trial. Since trials rarely fulfil both of these conditions, pragmatic trials usually borrow from the left-hand column of *Table 7* by specifying the significance tests that they will undertake.

These far-sighted distinctions influenced the pragmatic design of the COGNATE trial in at least four ways:

1. While fastidious trials mimic the laboratory conditions associated with scientific investigation, pragmatic trials take place in normal clinical practice.

TABLE 7 Differences between pragmatic and fastidious trials

Topic	Type of trial	
	Fastidious	Pragmatic
Objective	To acquire information relevant to defined scientific hypotheses and, thus, to draw scientific conclusions	To decide between two treatments in clinical practice rather than under ideal conditions
Definition of treatment	Rigid and equalised; in particular the trial protocol defines treatments so that psychosomatic or placebo effects are the same for each treatment	Flexible and optimal; in particular the protocol defines treatments so that each makes the best of psychosomatic or placebo effects
Experimental conditions	Tightly controlled laboratory conditions	Normal clinical practice
Definition of patients	The trial protocol strictly defines those patients eligible for all trial treatments prospectively, but may revise that definition retrospectively. Patients who withdraw from allocated treatments thereby withdraw from the analysis	The trial protocol defines patients eligible for all trial treatments flexibly but irrevocably once randomisation has occurred. Patients who withdraw from their allocated treatments after randomisation remain in the analysis
Number of criteria	No constraint on the number of criteria provided the trial protocol defines all criteria in advance	Only one; hence, if there are many potential criteria, the trial protocol must give them empirical weights so as to yield a single decision function, for example cost per QALY
Method of analysis	Traditional significance test for each hypothesis (but no formal relationship between significance tests)	Select treatment which gives the best weighted decision function (but no formal significance test)

QALY, quality-adjusted life-year.
Derived from Schwartz and Lellouch 1967.[109]

2. While fastidious trials define treatments rigidly, so as to keep hypothesis tests free from external influence, pragmatic trials define treatments flexibly because they seek the best decision for the complexities of normal clinical practice.
3. While fastidious trials seek to equalise placebo or non-specific effects, so as to compare like strictly with like, pragmatic trials seek to optimise these effects as one does in clinical practice
4. While fastidious trials exclude from analysis participants who violate the protocol in any way ('analysis per protocol'), pragmatic trials seek to analyse all participants according to their allocated treatment whatever subsequently happens ('analysis by treatment allocated').[110]

The value of EUS in staging patients with gastro-oesophageal cancer is not proven. Hence the only ethical means of evaluating this investigation is a randomised trial. As the funders – the National Institute for Health Research (NIHR) – aim to inform decision-making in the NHS, and EUS was already widespread across the NHS, the trial has to be pragmatic. Accepting these arguments, the Multicentre Research Ethics Committee (MREC) for Scotland approved this pragmatic protocol on 14 June 2004. Thus the scientific validity of the COGNATE trial lies in its adherence to the pragmatic scientific paradigm rather than the fastidious scientific paradigm.

It is intrinsic in the pragmatic scientific paradigm that, after randomisation between alternative interventions (in this trial, alternative diagnostic pathways and their therapeutic sequelae), multidisciplinary teams (MDTs) and individual clinicians make optimal clinical decisions for trial participants. Thus the COGNATE trial is evaluating, not an isolated EUS scan seen as a simple diagnostic intervention, but the 'complex intervention'[111] comprising the entire sequence of clinical decisions that flow from that intervention. In particular, the COGNATE economists seek to cost all the consequences for the use of NHS resources that lie downstream from the focal EUS scan or its absence.

Quality assurance

Little guidance is available for assuring the quality of the clinical processes in clinical trials. Most trials, including COGNATE, follow standard operating procedures (SOPs)[112] providing rigorous guidance on the conduct of the trial itself. However there is little if any scientific literature on ensuring the quality of the clinical process that is being tested. As EUS is an operator-dependent skill, it was important to assess the quality of the scanning process within the COGNATE trial.

Variation in the interpretation of scans has three main sources. First there are concerns over the accuracy of EUS scans[36] and the learning curve of those who interpret them.[113] Secondly the equipment to record and store images is not consistent between centres. Thirdly analytical interpretation of scans varies among observers and even over time by the same observer. Hence the COGNATE trial aimed to develop and report on a prospective system of peer review to assure the quality of EUS scans. In particular, it reviewed the quality of the reports and recommendations made by reporting clinicians.

Chapter 2 Methods

Design

We conducted a pragmatic multicentre randomised trial to evaluate the (clinical) effectiveness and cost-effectiveness of EUS as a technology to improve the staging and thus the management of gastro-oesophageal cancer. In planning the trial, we assessed tools for estimating quality of life in patients with gastro-oesophageal cancer. The ensuing psychometric analysis of data collected at baseline and after 1 and 3 months enabled us to develop an appropriate outcome measure for quality of life, which we used in the effectiveness and cost-effectiveness analyses. To ensure that the trial evaluated 'current best practice' in endosonography, we initiated a rigorous quality assurance process.

Intervention

We designed the COGNATE trial to test the effect on quality-adjusted survival of undergoing EUS within the staging process. Before the trial began, we developed a pragmatic, and therefore advisory, staging algorithm from normal clinical practice as characterised by SAGOC:[4]

1. All patients should receive biochemistry, haematology, pulmonary function tests and cardiac assessment, not least to exclude patients whose World Health Organization (WHO) status is 3 or 4, or who are medically unsuitable for either surgery or chemotherapy.
2. Patients who are medically fit for surgery without evidence of metastases should undergo CT following an agreed protocol using spiral scanner and intravenous contrast.
3. Patients with any suspicion of peritoneal disease should undergo laparoscopy as the best means of detecting peritoneal tumour deposits.
4. Fit patients with localised tumours and no contraindications were eligible for randomisation to EUS or not.

In the resulting control group (or 'non-EUS group' in tables), the choice of treatment depended on the results of the completed initial staging investigations, revisited if necessary. In the resulting intervention group (or 'EUS group' in tables), the final choice of treatment followed the EUS scan. At the end of staging, with or without EUS, MDTs assigned patients to one of three treatment options. Patients with:

1. tumours that were adjudged to be mucosal underwent EMR with or without argon-beam ablation of the surrounding mucosa
2. tumours that were adjudged to be resectable underwent surgery with or without neo-adjuvant chemotherapy, typically with cisplatin and 5FU
3. advanced localised disease, for which complete resection was adjudged to be impossible, received multimodal treatment, possibly including palliative surgery for gastric cancers.

Thus we randomised no patients who then had evidence of metastases or then had plans for palliative treatment or were then known to be medically unfit for surgery. In a pragmatic trial, of course, subsequent changes in all this information cannot invalidate the randomisation.

Trial flow chart

Figures 1 and *2* summarise the trial design. Randomisation took place after review of the initial staging investigations by the MDT. Clinicians agreed a conditional management plan, sought informed consent and randomised patients between receiving EUS and proceeding to the agreed management plan. They also reported to the North Wales Organisation for Randomised Trials in Health (NWORTH), Bangor University's Registered Clinical Trial Unit, all patients whom the MDT decided not to randomise
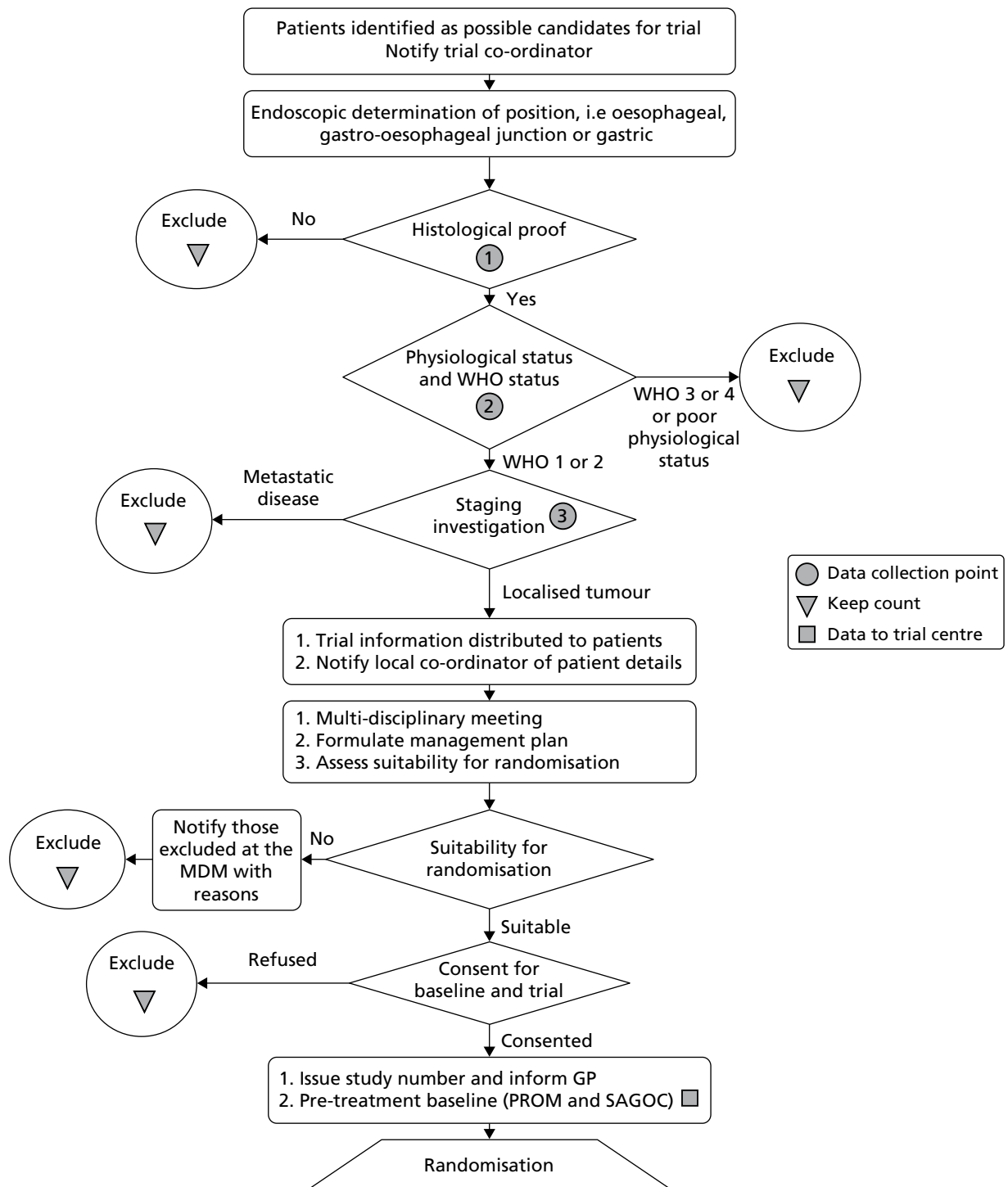
FIGURE 1 Trial design: start to randomisation. MDM, multidisciplinary meeting. PROM, patiented-reported outcome measure.

with reasons for exclusion. These included patients for whom they considered EUS either essential or inappropriate.

Inclusion and exclusion criteria

The COGNATE trial was a trial of patients with proven cancer of the oesophagus, stomach or gastro-oesophageal junction. To be eligible for the trial, patients had to be fit for both surgery

```
                              _____
                             /  Randomisation  \
                            /_____\
                          /                        \
                        /                            \
            ┌──────────────────┐          ┌──────────────────────────┐
            │    EUS group     │          │  Treatment as usual group │
            └──────────────────┘          └──────────────────────────┘
                     │                                  │
                     ▼                                  ▼
      ┌────────────────────────────┐    ┌────────────────────────────┐
      │ Patients assigned to one of:│    │ Patients assigned to one of:│
      │ 1. Multimodal treatment     │    │ 1. Multimodal treatment     │
      │ 2. Endoscopic mucosal       │    │ 2. Endoscopic mucosal       │
      │    reduction                │    │    reduction                │
      │    ± mucosal ablation       │    │    ± mucosal ablation       │
      │ 3. Surgery ± neo-adjuvant   │    │ 3. Surgery ± neo-adjuvant   │
      │    therapy                  │    │    therapy                  │
      └────────────────────────────┘    └────────────────────────────┘
                     \                          /
                      \                        /
                       ▼                      ▼
              ┌──────────────────────────────────┐        ┌──────────────────────┐
              │ Discharge notification        ▣  │        │ ▣ Data to trial centre│
              │ 1. Treatment received            │        └──────────────────────┘
              │ 2. Morbidity/mortality           │
              │ 3. Completeness of resection     │
              └──────────────────────────────────┘
                             │
                             ▼
              ┌──────────────────────────────────┐
              │ 1-month follow-up            ▣   │
              │ 1. Survival                       │
              │ 2. Disease recurrence             │
              │ 3. PROM and some SAGOC            │
              └──────────────────────────────────┘
                             │
                             ▼
              ┌──────────────────────────────────┐
              │ 3-, 6-, 12-, 18-, 24- and         │
              │ 36-month follow-up          ▣    │
              │ 1. Survival                       │
              │ 2. Disease recurrence             │
              │ 3. PROM and some SAGOC            │
              └──────────────────────────────────┘
                             │
                             ▼
              ┌──────────────────────────────────┐
              │      Final follow-up   ▣         │
              └──────────────────────────────────┘
```

FIGURE 2 Trial design: randomisation to conclusion. PROM, patient-reported outcome measure.

and chemoradiotherapy as well as free of metastatic disease. Both their ASA (American Society of Anesthesiologists) grading[114] and their WHO performance status[115] had to be 1 or 2 (see *Figure 1*). Following initial staging, clinicians could exclude patients from the trial for clinical reasons.

Patient information and informed consent

Before randomisation, the research professionals or clinicians invited eligible patients to participate in the COGNATE trial, gave them the patient information sheet approved by the Scotland MREC, and allowed them time to consider it and ask questions. We explained the nature of EUS and the process of randomisation to these patients. We stressed that the choice of treatment after EUS was the same in both groups. We then asked consenting patients to sign the consent form.

Randomisation

Once an eligible patient had consented and completed the baseline quality-of-life questionnaire, the recruiting centre telephoned NWORTH in Bangor. NWORTH staff confirmed eligibility and asked for information on both stratifying variables: centre and tumour location – gastric, oesophageal or the gastro-oesophageal junction. As we included only participants with good WHO performance status, we did not need to stratify for this. NWORTH then randomised the participant between intervention and control

groups using a dynamic randomisation algorithm designed to prevent subversion.[116,117] We reported regularly on recruitment to the National Clinical Research Network and NCCHTA. NWORTH confirmed the allocation by e-mail to the recruiting centre, which either booked an EUS scan for intervention participants or continued the agreed management plan for control participants.

Sample size

Our original application proposed to consent, randomise and follow up a total of 700 patients in a trial in which the primary outcome was survival. We soon discovered that most centres in the UK wanted to use EUS to stage gastro-oesophageal cancer. Conscious that early participants were able and happy to report on their health-related quality of life (HRQoL), however, we decided with the support of the Trial Steering Committee (TSC) and the approval of the Health Technology Assessment (HTA) programme to change the primary outcome to quality-adjusted survival. This effectively combines the components of health that EUS might improve – survival through better staging and HRQoL through reassurance arising from better staging and planning – in principle reducing the sample size needed.

As there is no easy means of calculating the power of a sample for the primary outcome of quality-adjusted survival, we calculated power for two simple but plausible scenarios. First, if there were no difference between groups in quality of life, the combination of a sample of 400 participants and a log-rank test using a 5% significance level would yield 80% power of detecting a hazard ratio of 0.6, equivalent to a difference between 60% survival at 12 months [derived from SAGOC:[4] *Appendix 1* and *Figure 2*] and 73% survival. Second, if there were no difference between groups in survival, the combination of the sample of 400 and a *t*-test using a 5% significance level would yield more than 80% power of detecting a 'small' effect size of 0.3[118] in quality of life. As the groups were more likely to differ in both survival and quality of life, the power of our primary analysis of quality-adjusted survival would be correspondingly greater. At worst, if we were able to randomise and follow-up only 220 patients, that scenario would yield 80% power to detect a hazard ratio of 0.5 (equivalent to a difference between 60% and 78% in survival at 12 months) or an effect size of 0.4 in quality of life, still 'small'.[119]

Quality-of-life instruments

The COGNATE trial used two instruments to gather information on quality of life as the basis for evaluating both effectiveness and cost-effectiveness: the European Quality of Life – 5 Dimensions (EQ-5D) and its visual analogue scale (EQ-VAS); and the Functional Assessment of Cancer Therapy (FACT) comprising FACT – General (FACT-G) and FACT – Additional Concerns (FACT-AC). Centres administered the questionnaires, including these instruments at baseline and 1, 3, 6 and 12 months after randomisation and at 18, 24 and 36 months where possible.

We used the EQ-5D, developed by the EuroQol Group,[120,121] to measure patients' HRQoL and to ascribe utilities to their health states. The EQ-5D is a preference-based generic measure comprising five domains: mobility, self-care, usual activities, pain and discomfort, and anxiety and depression. Each domain has three levels: no problems, some problems and a lot of problems. The EQ-5D scoring system defines 245 possible health states, namely $3 \times 3 \times 3 \times 3 \times 3$, plus two additional states – dead and unconscious. EQ-5D gives death a utility of zero and 'best imaginable health' a utility of one. For each participant it converts the five item scores into a summary utility based on the 'time trade-off' preferences of a UK-wide random sample of 3000 respondents.[122] Some health states have negative utility ('worse than death'). We included the EQ-5D in our outcome portfolio as the primary means of adjusting survival for quality of life. EQ-5D complements its five items with a visual analogue scale (VAS); this is a single thermometer-like generic quality-of-life scale with zero representing 'worst imaginable health' and 100 representing 'best imaginable health', on which respondents mark their perceived current health directly. Scores, therefore, require no further processing.

Functional Assessment of Cancer Therapy is a psychometric instrument measuring cancer-specific quality of life.[123,124] The current version of FACT-G comprises four subscales: Physical Well-Being (seven items), Social

or Family Well-Being (seven items), Emotional Well-Being (six items) and Functional Well-Being (seven items). FACT sums scores on these subscales to give FACT-G. The COGNATE team derived its FACT-AC scale from Gastric Additional Concerns,[125] comprising 19 items, and Oesophageal Additional Concerns,[126] comprising 17 items, by removing overlapping items and psychometrically weak items using methods described by Streiner and Norman.[119] In this way we effectively merged the Gastric Additional Concerns and Oesophageal Additional Concerns scales to form a single integrated Gastro-Oesophageal Concerns scale for easy use by all trial patients. We also assessed the extent to which this provided information over and above that provided by EQ-5D.

Data collection

Centres collected data on the due day when possible, but otherwise within a window. Although pre-randomisation data could be collected up to 3 days before randomisation, randomisation could not proceed without these data. Data due at 1, 3 and 6 months could be collected up to 14 days after the due date; data due at 12, 18, 24 and 36 months could be collected up to 28 days after the due date. To avoid bias, for example by anticipating a major event, we did not collect data before the due date except for pre-randomisation data. Similarly, we did not collect data for 7 days after a major procedure.

Our preferred mode of administration was a rigorously defined face-to-face interview. In our experience interviews reduce biases due to sicker patients not responding. Trained research professionals read each question while participants followed laminated versions. The researchers entered their responses directly into a laptop computer. If a face-to-face interview was not possible, we conducted the interview by telephone, having posted the questionnaire to the participant in advance. As a last resort, the participant could complete the questionnaire and return it by post; with that exception, research nurses were interviewers not observers. The researchers also recorded the mode of completion: face to face, telephone or postal. Although we know that interviewers affect responses to questionnaires, there is strong evidence [from a recent systematic review to which two of us (DKI, ITR) contributed] that this effect is consistent across trial arms.[127] Furthermore FACT itself is robust against interviewer effects.[128]

Primary outcome measure: quality-adjusted survival

The primary outcome measure was quality-adjusted survival, using the EQ-5D health index to adjust for the quality of life of survivors. We integrated outcomes over time for individuals by calculating the 'area under the curve' (AUC). This avoids multiple testing of correlated outcomes. We calculated this area from the graph of HRQoL (EQ-5D or FACT) against time by joining all the intermediate points derived from the follow-up interviews, drawing vertical lines to the horizontal (time) axis at randomisation and at death, complete withdrawal or censoring at the end of the trial (August 2009), whichever occurred soonest, and then calculating the area of the resulting polygon. This area is the standard measure of quality-adjusted life-years (QALYs) in cost–utility analysis.[129]

Secondary outcome measures

Survival adjusted by FACT
It is possible to use quality-of-life measures other than EQ-5D to adjust survival for quality of life. We used AUC summaries of the two main FACT scores, FACT-G and the combined Additional Concerns (FACT-AC), as cancer-specific and site-specific versions of quality-adjusted survival. We converted both measures to a scale with minimum 0 (worst quality of life) and maximum 1 (best quality of life) before calculating the AUC.

Survival from randomisation
However standard survival analysis uses only available information on participants' survival, including those withdrawing from the trial, and takes account of variable follow-up by censoring observations. Hence no imputation is necessary. The maximum observation time was 12 months for those last randomised and 58 months for those first randomised to the pilot study.

Quality of life at 12 months
We compared all three measures – EQ-5D, FACT-G and FACT-AC – between intervention and control groups at the 12-month interview. This was the minimum planned length of follow-up between randomisation and the end of the trial. Data from later interviews could be compared only on subsamples recruited nearer the start of the trial.

Management plan changes
We recorded changes to the management plan agreed by the MDT that occurred after randomisation and the treatment actually received.

Quality of treatment
1. Complete resection, defined as the removal of all macroscopic and microscopic lesions, was recorded by participating centres for the whole sample on the trial database. For the few participants for whom this conclusion was missing, the three clinical members of the trial team independently assessed all other information about excision and reached consensus while blind to allocated treatment.
2. Pathological reporting of resected specimens using SAGOC criteria.[4]
3. Morbidity and mortality potentially caused by EUS. We asked participating centres to record all complications of EUS (including deaths in hospital within 30 days) that might have been related to the investigation. There were no such complications. Indeed the only early death of an intervention participant occurred after palliative surgery without any suspicion that EUS played any role.

Cancer mortality according to the SAGOC definitions
For all deaths we classified the cause as EUS related, completely or partly cancer related, related to cancer treatment, or unrelated to any of these. We also recorded all diagnoses of metastases, either during life or at death.

Follow-up
We followed patients until death or the end of data collection, which was between 12 and 58 months after the end of recruitment for all patients. We collected data at discharge from hospital after initial treatment and at follow-up clinics after 1, 3, 6, 12, 18, 24 and 36 months.

Quality assurance of endoscopic ultrasound scans
The COGNATE team asked investigators to record EUS scans on videotapes and to record staging and explanatory comments on a trial proforma. Anonymised videos of selected EUS scans of intervention participants were reviewed by a panel of investigators during a series of web conferences, each of which reached a blinded consensus on the staging of four tumours. We compared the staging decisions of the individual peer reviewers, their consensus decision and the staging decision of an international expert as external reviewer with that of the original investigator who performed the scan. We used Cohen's kappa or weighted kappa[130] to test for agreement.

Independent trial monitoring
The COGNATE trial had a TSC comprising an independent chairperson, two independent members and four members of the COGNATE Trial Executive Group (TEG). Principal investigators (PIs) in each centre reported to the TSC through the TEG. The Data Monitoring and Ethics Committee (DMEC), comprising three independent members with the trial statistician in attendance, acted as a subcommittee of the TSC, reporting to the TSC. *Appendix 3* lists members of the TSC and DMEC.

Trial management
North Wales Organisation for Randomised Trials in Health managed the trial. A Trial Management Group comprising the two chief investigators, trial statistician, health economist, outcomes specialist, data manager, trial manager and trial administrator met monthly at NWORTH. Telephone access was available to other investigators by invitation. Minutes were taken and stored in the Trial Master File.

Ethics and participating sites

The COGNATE trial was approved by the Scotland MREC A and 10 Local Research Ethics Committees (LRECs) and associated research governance units for a total of 16 hospital sites. Two centres did not recruit any patients, one owing to competing workload and the other (who had obtained ethical approval from four LRECs for seven hospitals in that area) owing to lack of commitment from staff. *Appendix 2* gives details of the eight recruiting hospitals.

One centre screened many patients but randomised only one into the COGNATE trial, and later asked to withdraw from the study. They agreed that their COGNATE patient, if willing, could receive continuing follow-up through a nurse from the local Cancer Research Network. This arrangement was accepted by the patient and the local Research Governance Department.

Protocol amendment

The proposal to change the primary outcome measure from survival to the more sensitive measure of quality-adjusted survival, which reduced the target sample size from 700 to a maximum of 400 and a minimum of 220, was agreed by the DMEC and TSC in April 2006 and reported to the NCCHTA in December 2006. A 6-month, no-cost extension was agreed by the DMEC and TSC in April 2007 and by the NCCHTA in November 2007. The final protocol is in *Appendix 1*.

Electronic data collection and storage

All data identified patients only by a unique trial number. Each trial centre kept its own index linking trial numbers to patients' names and addresses separate from the laptop computers used to store and transfer trial data, and protected that index by key and password. The trial co-ordinating centre in Bangor had no access to these local indices. Those analysing data had no access to which group of participants received the intervention until the TSC had scrutinised and approved the methods and findings of the main statistical analysis.

North Wales Organisation for Randomised Trials in Health designed a Microsoft Access database (Microsoft Corporation, Redmond, WA, USA) to enable each centre to collect COGNATE data on their trial laptop computer. The electronic forms for data entry were designed to be friendly to local staff. When local staff suggested an improvement to display the dates of tests already entered into the database, the database designer programmed the change and the trial manager implemented it on the next suitable occasion – monitoring visit, investigators' meeting or over the telephone.

Centres transferred data to NWORTH by electronic file transfer protocols every 2 months throughout the trial. This allowed the data manager and trial manager to monitor data completeness and report to the research team. As Bangor University updated file transfer protocols twice during the COGNATE trial, this necessitated retraining of local staff in methods of data transfer. By all of these means the COGNATE trial can fairly claim to have been a generally paperless trial.

Statistical methods

Statistical analysis plan

To avoid bias, we wrote our statistical analysis plan, and the TSC approved it, during the recruitment phase. Although participants and their clinicians knew their randomised allocation, we kept the methodological chief investigator and trial statistician blind to all these allocations until they had presented blinded findings to the TSC in September 2009.

Primary analysis was by 'whether allocated to endoscopic ultrasound'.[109,110] This reflects the pragmatic nature of the trial, and its primary goal of evaluating that health technology in informing decisions in the real world. We also undertook secondary analysis by 'endoscopic ultrasound received' to explore the implications of clinical decisions that diverged from the random allocation.

Quality-adjusted survival (primary outcome) and survival

The primary outcome measure (survival adjusted by EQ-5D), and survival-based secondary outcome measures (unadjusted survival, survival adjusted by FACT-G, and survival adjusted by FACT-AC) all analysed the time between randomisation and the end of the trial or death (if it occurred before the end of the trial). Thus surviving participants recruited at the end of the trial were censored at 12 months, while those recruited near the beginning of the trial were censored only if they survived 48 months or more.

Because of the variable follow-up, applying standard analysis of variance methods or *t*-tests to these measures is biased against the group with better survival. There are a range of survival analysis methods, all of which allow for censored observations and thus avoid this bias. For descriptive comparisons, we used Kaplan–Meier estimates of mean and median length of survival as numerical summaries, survival curves as graphical summaries, and log-rank tests to compare the survival experience of different groups. However we used Cox regression, which models the simultaneous effect on survival of several characteristics, for the main (quality-adjusted) survival comparisons.

Both primary quality-adjusted survival analysis and secondary survival analyses using Cox Regression models considered several baseline characteristics, including EQ-5D and FACT baseline scores, for inclusion as covariates. We did this, not only to take account of any baseline imbalance between groups despite stratification, but also to improve the precision and generalisability of the model. We always included centre, condensed to three groups of similar size: Aberdeen, Gloucester and the rest. We always used the baseline score of a given measure to predict a later score of that measure. Other characteristics considered in step-wise model building included: age and gender; site, stage and type of tumour; the initial management plan agreed before randomisation; but not WHO status, as most participants had a WHO status of 1. To get the best from 'initial management plan' in predicting outcomes, we created a binary variable to distinguish between conservative prior plans (namely chemotherapy, radiotherapy, both or neither) and therapeutic prior plans (namely endoscopic resection or surgery in some form). As we had expected, this later proved very good at predicting outcomes. As conservative plans choose between all possible combinations of chemotherapy and radiotherapy, we followed the example of many MDTs by describing this and the resulting binary variable as 'multimodal'.

Quality of life at 12 months: sensitivity analysis of primary outcome

We analysed EQ-5D, FACT-G and FACT-AC scores at 12 months by general linear models, using the same approach to covariates as in survival analyses. Although all follow-up times contributed to the EQ-5D- or FACT-adjusted survival analyses, we reported these three measures, and the four FACT subscales, at all follow-up times only descriptively to reduce multiple testing. As in the Cox regressions, we supplemented model-based analyses with basic summaries and graphs.

For cost-effectiveness analysis, we extended the primary QALYs measure beyond the end of the trial to 48 months for all participants by a combination of survival modelling and imputation. When all censored individuals have the same follow-up time, survival analysis methods are no longer needed. Hence this fully imputed measure, and the equivalent truncated to the minimum follow-up of 12 months, were analysed in the same way as the quality-of-life measures at 12 months. This provides both a sensitivity analysis for the main effectiveness measure and a link to the cost-effectiveness analysis.

Other outcome measures

We compared changes in management plan, complete resection rates and cause of death between groups and centres by chi-squared tests of the appropriate proportions. In response to referees, we also investigated the relationship between changes in management plans and selected other outcomes (survival, quality of life, cost and complete resection) for the two allocated groups.

Modelling: covariates and interactions

We used analysis of covariance (ANCOVA) in general, and Cox regression models specifically for survival and quality-adjusted survival, to improve comparisons between intervention and control groups. These

techniques enhance the basic analytical techniques – *t*-tests and log-rank tests respectively – by making allowance for covariates, namely participant characteristics that affect outcome. Most covariates have similar effects on intervention and control groups, therefore including them in models corrects for baseline imbalances if present, and thus improves estimates of the effect of EUS. Whether or not there are any baseline imbalances, covariates can also explain some of the intrinsic uncertainty and thus reduce it.

In choosing covariates we followed the analysis plan that we had defined prospectively and the independent TSC had approved, also prospectively. This plan considered as potential covariates only the pre-randomisation variables listed above [see *Quality-adjusted survival (primary outcome) and survival*]. First we entered centre and baseline values of the outcome under analysis. Secondly we sought covariates that affect that outcome directly; to avoid subjective choices, we used a step-wise procedure and included only covariates that increased the precision of the estimated effect of EUS.

However covariates may also interact with the effect of EUS in the sense that changing the value of the covariate changes the estimated difference between an intervention participant and a control participant with the same characteristics. Thus an interaction indicates which participants derive the most benefit from EUS. The first two analytical steps we have just described used covariates without interactions. The third step added interactions between the treatment allocated and the covariates chosen in the previous two steps if they improved the model significantly. In the only change to this analysis plan after the TSC had approved it, we also investigated whether or not the unexpected interaction between the effect of EUS on survival adjusted by EQ-5D and the baseline EQ-5D also applied to survival adjusted by FACT and the baseline EQ-5D.

Imputation and missing values

To avoid bias in analysis by treatment allocated, it is desirable not to exclude participants for whom some outcome data are missing;[129,131] whenever possible, therefore, we imputed these data from known data about these participants and other participants whose outcome data are known.[132,133]

The main trial recruited for 3.5 years. To maximise statistical power, therefore, we followed all participants for as long as possible, namely between 1 and 4.5 years. The main aim of imputation for the effectiveness analysis was to achieve complete data within this design rather than to extend data beyond 31 July 2009, the end of the trial and thus the censoring date. However the bootstrap methods used to assess cost-effectiveness required us to estimate costs and benefits for all participants for the same period. We chose two follow-up times: 12 months, the minimum unless a participant withdrew from the trial; and 48 months, which took account of all information on both survival (as there were no subsequent deaths) and quality of life (as the last questionnaire to participants was at 36 months). Sensitivity analyses of both effectiveness and cost-effectiveness provide explicit links between the effectiveness and cost-effectiveness sections of this report.

Potentially there was complete information for all participants who died during follow-up. Dead participants did not use any more resources, and we set their subsequent quality of life to zero, rather than missing. We also received survival data on all participants up to complete withdrawal or the end of the trial. As survival analysis allows for variable follow-up, it does not need imputation.

Data were of two types:

(a) *Clinical data, including demographic and resource use, abstracted by research professionals from hospital notes and entered retrospectively onto the electronic database*. In general we did not need to impute these data, because we asked research professionals to collect complete data except for pre-randomisation tumour stage, for which we permitted 'missing'. However we imputed resource use in secondary care from the end of the trial to 48 months.

(b) *Patient-reported outcome measures at baseline and follow-up*. We imputed the few missing data for the main effectiveness analysis. For the cost-effectiveness analysis and effectiveness sensitivity analysis, we needed to impute quality of life to 48 months.

We imputed missing quality-of-life data in three phases. Each phase used SPSS (SPSS Inc., Chicago, IL, USA) Missing Values Analysis (MVA) procedure,[132] which simultaneously estimates all missing values in a data set, on one or more data sets. The final phase also used estimated survival probabilities from the Cox regression model that we have described.

Imputing quality of life: phase 1 – psychometric and effectiveness analyses

Phase 1 used all quality-of-life items (27 in FACT-G covering four subscales; 33 in FACT-AC; five in EQ-5D, one for each domain; and EQ-VAS, a single item) answered in interviews at the same time to estimate the missing items in those interviews. This yielded a complete set of responses for existing interviews. Initial psychometric analyses, using responses at 0, 1 and 3 months, reduced the number of items in FACT-AC by two, after which we repeated phase 1. We used the resulting data to calculate scores for EQ-5D and FACT scales and subscales for all existing interviews.

Imputing quality of life: phase 2 – until 12 months

We then discarded item scores, except EQ-VAS, in favour of scale scores across time. We created a single data set comprising all 213 participants and set scale scores to zero after death. To this data set we added the allocated treatment, and baseline characteristics to improve estimates. Phase 2 used only time points up to 12 months, the minimum period of follow-up in the trial. We used MVA to impute scale scores at times without interviews for those who were still alive at 12 months, and then for those who had died by 12 months. This yielded a complete imputed data set with all quality-of-life scores at all times up to 12 months. Three participants withdrew before 12 months. While phase 2 included them among survivors, phase 3 adjusted their estimated quality of life at times after withdrawal to take account of the probability of death.

Imputing quality of life: phase 3

No more participants withdrew completely after 12 months. Beyond 12 months, however, the survival status of progressively more participants is unknown because of censoring at the end of the trial. Phase 3 therefore estimated both the probability of being alive at each of the three remaining times and the quality of life of the participant if alive. Multiplying these two estimates yields the expected quality of life. This procedure adapts to quality-of-life data the process for imputing censored cost data described by Lin *et al*.[133]

In the first part of phase 3, we derived the probability of censored participants being alive at 18 months from the Cox regression model for survival. We calculated similar probabilities at 24 and 36 months, and also at 48 months for use in cost-effectiveness analysis and effectiveness sensitivity analysis. In the second part of phase 3, we used three separate MVA imputations to extend the data set from phase 2 to 18 months, 24 months and 36 months for those not known to be dead at those times. We multiplied each imputation by the probability that each participant would have been alive at this time. Finally we set quality-of-life scores for people known to be dead to zero, or the equivalent for FACT-AC, for which 0 is the best possible score.

By the end of phase 3, we had complete quality-of-life information for all 213 participants at all time points before the end of the trial, and survival status at that date, enabling us to estimate QALYs for primary analysis. We also had expected quality-of-life scores, but not survival status, for all 213 participants at all time points up to 36 months, for cost-effectiveness analysis and fully imputed sensitivity analysis for effectiveness.

Imputing: phase 4 – secondary care costs

We combined data on resource use in secondary care into six periods – up to 12 months, 12–18 months, 18–24 months, 24–36 months, 36–48 months and 48–60 months. We costed and summed unimputed frequency data to give the total cost in each period for each participant. However we were able to discard the final period because by 48 months only 10 participants were still in the trial, none of whom reached 60 months. For each of the first five periods we imputed the expected cost for unobserved participants by adapting the method of Lin *et al.*,[133] although not exactly as we had adapted it to quality-of-life data in phase 3 above.

Of these four phases of data imputation, phase 1, which imputes missing answers to questions within a scale from answers to related questions in the same scale, is not appropriate to costs because costs have no 'related questions within a scale' in the psychometric sense. Phase 2 was not necessary because we observed costs until censoring at 12 months or later. As with Lin *et al.*,[133] our costs were spread over intervals, while we had collected and imputed quality-of-life scores at exact time points, which included the ends of the cost intervals. Hence, although the survival probabilities were exactly the same in Phases 3 and 4, we used two for each cost interval – those of being alive at the start and the end of the interval. Unlike quality-of-life scores, however, costs are highly skewed and unsuitable for the MVA procedure.[129] In general, therefore, we estimated costs in unobserved intervals from the mean cost among people in the same allocated treatment group who were alive and observed throughout the interval. Nevertheless we used separate estimates for the cost of the year before death, because they are consistently and considerably higher than all years other than the first.

Psychometric methods to refine quality-of-life assessment

Rationale

We assessed quality of life by the EuroQol[120,121] and FACT[123,124] (see *Design*, *Quality of life instruments*). We chose the EQ-5D as the primary means of adjusting survival for quality of life. We included the EQ-VAS as a natural adjunct to the EQ-5D. We added FACT as an alternative means of adjusting survival, conditional on psychometric analysis within the trial. When the primary outcome measure changed from survival to quality-adjusted survival during the trial, thorough quality-of-life assessment became even more crucial.

The standard way to score FACT data is as follows. Each subscale is the sum of its items. FACT-G is the sum of Physical, Social, Emotional and Functional subscales. FACT Total is the sum of FACT-G and Additional Concerns. FACT Trial Outcome Index (TOI) is the sum of Physical, Functional and Additional Concerns. The rationale for the TOI is:

> *It is a common endpoint used in clinical trials, because it is responsive to change in physical or functional outcomes. While social and emotional well-being are very important to quality of life, they are not as likely to change as quickly over time or in response to therapy.*[134]

To assess the best use of FACT in the COGNATE trial, we posed three research questions and conducted psychometric analyses separately on the data collected before randomisation, and at 1 and 3 months, when enough individuals attempted the questionnaires to make the analyses feasible.

Questions and corresponding analyses

How to score FACT-AC?

The COGNATE trial combined the two 'Additional Concerns' scales – Gastric and Oesophageal – so that all participants completed both scales. This could have resulted in some redundancy among the items measuring each construct. Although the scoring method for FACT-AC– summing the items – implies a single dimension, this has not been systematically tested. Therefore we decided to examine the factor

structure of the Additional Concerns items, and to consider the implications for scoring. We subjected the FACT-AC items to principal components analysis using SPSS version 16,[135] basing the number of components to be retained on the scree plot of eigenvalues and on the meaningfulness of the solution. We rotated these components orthogonally.

How to score FACT-G?

The scoring method of FACT-G – summing item scores to produce four subscale scores that can be summed into a single score – implies four intercorrelated dimensions. Previous factor analyses have typically found such a structure.[136] However we felt it prudent to check the factor structure of the FACT-G items and to consider the implications for scoring. We achieved this by subjecting the FACT-G items, that is to say the Physical, Functional, Social, and Emotional subscales, to confirmatory factor analyses with maximum likelihood estimation, using AMOS™ version 16[137] (SPSS Inc., Chicago, IL, USA). The factors were free to correlate. For each model, we considered the global fit, the residual covariances, the loadings of items on their intended factors, the modification indices for the loadings of items on their non-intended factors, and the correlations between factors. We measured global fit by the comparative fit index (CFI), root-mean-square error of approximation (RMSEA) and standardised root-mean-square residual (SRMR). Values of CFI above 0.9, SRMR below 0.10 and RMSEA below 0.08 represent adequate fit.[138] Ideally, we sought a significance level of at least 0.05 for the RMSEA test of close fit, where close fit is defined as an RMSEA value no greater than 0.05. We modified models only when there was both statistical and theoretical justification for so doing.

What FACT summary measure to use in COGNATE?

Although we included FACT as an alternative means of adjusting survival, it was not clear a priori which summary score would be best – General, Total or TOI. To make an informed choice, we examined relationships among the subscales and with the EuroQol measures using structural equation modelling with AMOS. We modelled the EQ-5D construct as an observed variable, and the FACT constructs as latent variables. It was not possible to use multiple indicators for these latent variables, given the size of the model and the size of the sample. Therefore we used an alternative means of adjusting for measurement error.[139] For each FACT latent variable there was one indicator, the corresponding FACT subscale. We fixed the path from the latent variable to the indicator at 1, and the measurement error at the variance of the indicator multiplied by 1 minus the reliability of the indicator. In the initial model, Additional Concerns was free to influence each of four FACT-G constructs, and each of the FACT-G constructs was free to influence EQ-5D, but the Additional Concerns construct was not free to influence EQ-5D directly. We allowed the disturbances of the four FACT-G constructs to co-vary. The criteria for adequate fit were as for the confirmatory factor analysis. We modified these models only when there was both statistical and theoretical justification for so doing. Finally we deleted non-significant paths and re-ran the model.

Health economic methods

Introduction

The primary objective of the health economic analysis was to assess the cost-effectiveness of EUS staging in the management and treatment of patients with gastro-oesophageal cancer compared with its absence. Our reporting of economic analysis in COGNATE is consistent with that in the Multi-Institution Nurse Endoscopy Trial (MINuET),[140] our published SOP for economic evaluation alongside RCTs,[141] and published guidance.[129,142]

Existing economic evidence

We found five relevant economic studies on staging cancer of the oesophagus.[143–147] Two were review articles[143,144] and three were American decision-analytic modelling studies.[145–147] Of these, Wallace et al.[147] is the most robust, arguing that CT, EUS and fine-needle aspiration (FNA) was the least costly strategy (US$40,000) and offered more QALYs on average (0.965) than all other strategies with the exception of PET, EUS and FNA (US$45,000 for 1.034 QALYs). Thus the latter was slightly more effective but also more

expensive, yielding a marginal cost-effectiveness ratio of US$61,000 per QALY, less than that of many medical treatments but above accepted thresholds in the USA and UK. We did not find any National Institute for Health and Care Excellence (NICE) guidance on this topic.

Outcome measures

The outcome measure for the economic analysis was the QALY. Both NICE and NCCHTA support the use of QALYs as an outcome measure in technology assessment.[140,148] QALYs measure health gain, combining survival and HRQoL.[149] The COGNATE trial measured HRQoL by the EQ-5D questionnaire.[120,121] We calculated the difference in mean QALYs between the intervention and control groups.

Analysis by time intervals

As participants were in the trial for different lengths of time, we adjusted estimates of costs and QALYs to allow for censoring. Following Lin *et al*.,[133] we used five time intervals covering 4 years (see *Statistical methods*). In the first interval (up to 12 months) both cost and QALY data were uncensored in principle. As we expected, the majority of costs fell during that first year. Hence we decided to conduct economic analysis over two different periods – for 12 months after randomisation and for 48 months, essentially the whole study period (see *Statistical methods*). In both we used bootstrapping, first to derive confidence intervals (CIs) around our point estimates of incremental cost-effectiveness ratios (ICERs) when appropriate, and then to draw cost-effectiveness acceptability curves (CEACs) to quantify uncertainty and convey to policy-makers the probability that EUS is cost-effective at different thresholds. In our primary analysis at 48 months, we discounted costs and QALYs at 3.5% per year.[148] We undertook sensitivity analyses at 12 and 48 months by varying the costs of EUS. We also re-ran our analysis using undiscounted QALYs at 48 months, again varying the costs of EUS. We undertook exploratory subgroup analyses at 48 months to test the effect of baseline EQ-5D and age on our results.

Resource use in secondary care

The COGNATE trial measured each participant's type and frequency of contacts with NHS secondary care over 4 years – contacts very likely to account for almost all their direct NHS costs. Within cancer care we focused on the EUS procedure, surgery, chemotherapy, radiotherapy, other drugs, inpatient stays, and day care and outpatient appointments. We integrated the collection of all these data on resource use into this essentially paperless trial. In particular we designed an electronic version of the Client Service Receipt Inventory (CSRI)[150,151] – a trial-specific structured form that enables research staff to report on the type and frequency of participants' contacts with health care. We took special care to check and maintain the quality of these data, unfamiliar to many trial practitioners. This led to a complete set of secondary care use data for the 213 participants who contributed to our analysis of clinical effectiveness and cost-effectiveness.

We excluded primary care costs for two reasons. First, the COGNATE team, including surgeons, economists and other health services researchers, were confident from our combined experience that these would account for a very small proportion of direct NHS costs. Second, although we initially tried to collect primary care data with quality-of-life data during structured interviews in secondary care, practical constraints led us to collect most of these primary care data by post. Subsequent rigorous validation established that the resulting primary care data, unlike the simpler quality-of-life data, were not good enough for robust imputation. After the unsuccessful validation, we followed the principles of Good Clinical Practice (GCP) by discarding the flawed primary care data, and changed the perspective of our economic analysis to secondary care.

As resource use is generally skewed, our published SOP for economic analysis advocates converting resources to costs, which are commensurable and facilitate analysis of uncertainty.[141] As distributions of resources and costs are especially skewed in cancer care, we always planned to impute censored data and bootstrap skewed data, rather than adopt the inappropriate method of adjusting observed costs by ANCOVA. Moreover, if COGNATE were to find significant differences in survival, and thus follow-up time, between allocated treatments, any simple comparison of observed costs would be biased.

Unit costs

Our general approach to costing all these data on resource use was to apply NHS reference costs for 2008[152] and drug costs from the *Prescription Cost Analysis 2008 England* (PCA).[153] We undertook detailed costing with oncologists, radiologists, surgeons and others at COGNATE sites for high-cost items such as multimodal treatment, neo-adjuvant therapy, surgery, chemotherapy, radiotherapy and positron emission tomography (PET) scanning. We derived the cost of a stent from Shenfine *et al.*,[154] and inflated it to 2008 prices using the Hospital and Community Health Services Pay and Price Index.[155]

We undertook detailed costing of the EUS procedure through a time-and-motion approach and discussion with oncologists, radiologists and surgeons at trial sites. We included variation between local and national estimates as part of our sensitivity analysis. *Table 8* summarises the sources of data on resource use and unit costs for secondary care.

TABLE 8 Sources of resource-use data and unit costs

Item of resource use	Source of resource use	Source of unit costs
Inpatient stay, day cases, outpatient appointments, treatments, etc.	Patient notes and prompted fields on laptop computer database	National *NHS Reference Costs 2007–8*[152]
Prescribing	Patient notes and prompted fields on laptop computer database	PCA[153]
Stent	Patient notes and prompted fields on laptop computer database	Shenfine 2005[154] – updated to 2008 prices
PET scan	Patient notes and prompted fields on laptop computer database	Aberdeen Royal Infirmary: Nuclear Medicine Department

Cost-effectiveness analysis

Our analysis of cost-effectiveness took an NHS perspective and compared effects by treatment allocated. As the trial randomised participants, we compared their costs between intervention and control groups. Cost-effectiveness analysis calculates ICERs, calculated as (C2–C1)/(E2–E1), where C2 is the mean total cost of the intervention group (EUS staging), C1 is the mean total cost of the control group (clinical management as usual), E2 is the mean effect in the intervention group and E1 is the mean effect in the control group. However interpretation of ICERs that cover more than one quadrant of the cost-effectiveness plane requires caution. In particular negative ICERs, which describe situations where one group (either intervention or control) is both less costly and more effective than the other, are not useful. As cost gains and effect gains are not competing but combining, their ratio is no longer relevant.

One alternative is the net monetary benefit (NMB) equal to the product of net QALYs and (WTP minus cost), where WTP is the decision-makers' maximum willingness to pay for a QALY.[129] The main disadvantages are that interpretation is difficult without separate knowledge of the costs and that the NMB for one condition cannot easily be related to NMBs for other conditions or to the NICE threshold. Instead we present CEACs, which combine the wider applicability of NMBs with a more user-friendly presentation. The CEAC is a graphical representation of the probability that an intervention is cost-effective over a range of monetary values for decision-makers' willingness to pay for a QALY. A value of zero for a QALY leads to comparison of the costs of EUS staging and usual care alone, and infinite or very large values result in a comparison of QALYs alone.

For COGNATE we use the values £20,000 and £30,000 – the range of threshold values used by NICE in the UK.[148] This enables policy-makers to compare the cost-effectiveness of EUS in gastro-oesophageal cancer with the range of estimated health gains per 'NHS pound' in other cancers and conditions.

We used bootstrapping to generate cost-effectiveness planes, CEACs and, where appropriate, 95% confidence intervals around our point estimates of ICERs.[156] Bootstrapping draws repeated samples from the trial data set with replacement. It calculates mean costs and effects in each group for each sample, and estimates the incremental cost-effectiveness of the intervention. We used Stata version 10.1 (StataCorp LP, College Station, TX, USA) to perform this non-parametric bootstrapping by generating 5000 samples.

Imputation, censoring, discounting and sensitivity analysis

In *Statistical methods, Imputation and missing values* we described the methods we used to impute missing cost and QALY data, and to allow for the censoring caused by variable follow-up in the COGNATE trial. We discounted all costs and QALYs beyond 12 months at 3.5% per year, as recommended by NICE.[148]

We conducted sensitivity analyses at both 12 months and 48 months to explore whether, and to what extent, the estimated ICERs for EUS relative to conventional staging are robust to key assumptions. There was uncertainty about the cost of EUS between local and national tariffs and estimates from the literature, so we explored the implications of the three different estimates we obtained. We also present undiscounted QALYs as another sensitivity analysis, not least for consistency with the effectiveness analysis.

Our analyses at 48 months include two subgroup analyses. First, we estimated CEACs for participants whose baseline EQ-5Ds were below or above the median score of the whole sample. Second, we estimated CEACs for participants whose ages at randomisation were below or above 65 years.

Quality assurance of endoscopic ultrasound scans

Objective

To develop a rigorous process of peer review to assure the quality of EUS scanning in a trial to evaluate the use of that technology in staging and managing gastro-oesophageal cancer.

Methods

The COGNATE site investigators reported the EUS scans of all participants in the intervention group by completing an anonymisable form (see *Appendix 8.1*) specifying tumour location and tumour, nodal and metastatic (TNM) stages. Our quality assurance panel comprised: an experienced surgical endosonographer and trialist as independent chair; six assessors from six sites – four radiologists, one gastroenterologist and one surgeon – who performed scans within the trial; with trial staff, notably the information technology specialist, in support.

The COGNATE site investigators recorded scans for peer review on to videotapes or compact discs (CDs) using trial number as sole identifier. They posted them by recorded delivery to the COGNATE team, who digitised and anonymised them, and stored them on the Bangor University website. The panel reviewed sampled scans at planned web conferences. One month before each conference we gave the panel links to sampled videos and asked them to study those videos on our website. When NHS firewalls denied access to the website, we posted them to members on CDs by recorded delivery. For each scan we asked members to complete and return a numbered assessment form (see *Appendix 8.2*) covering: the quality of the recording; their estimate of T, N and M stages; and the specific video times that informed those estimates. Trial staff collated results before each web conference.

Our IT specialist organised the conferences through www.webex.co.uk and transmitted videos and related documents to the office computers of all panel members. We recorded discussion both on audiotape and as text in the 'chat box' on the screen. If broadband problems affected the conference, the panel viewed CDs contemporaneously. Thus the panel viewed scans together but discussed staging blind, i.e. without knowing the source of the scan, the original assessment or those of fellow assessors. Once they had reached consensus on TNM stages, they could compare that consensus with the original assessment and review it if necessary. They also assessed the general quality of each scan and recording (*Figure 3*).
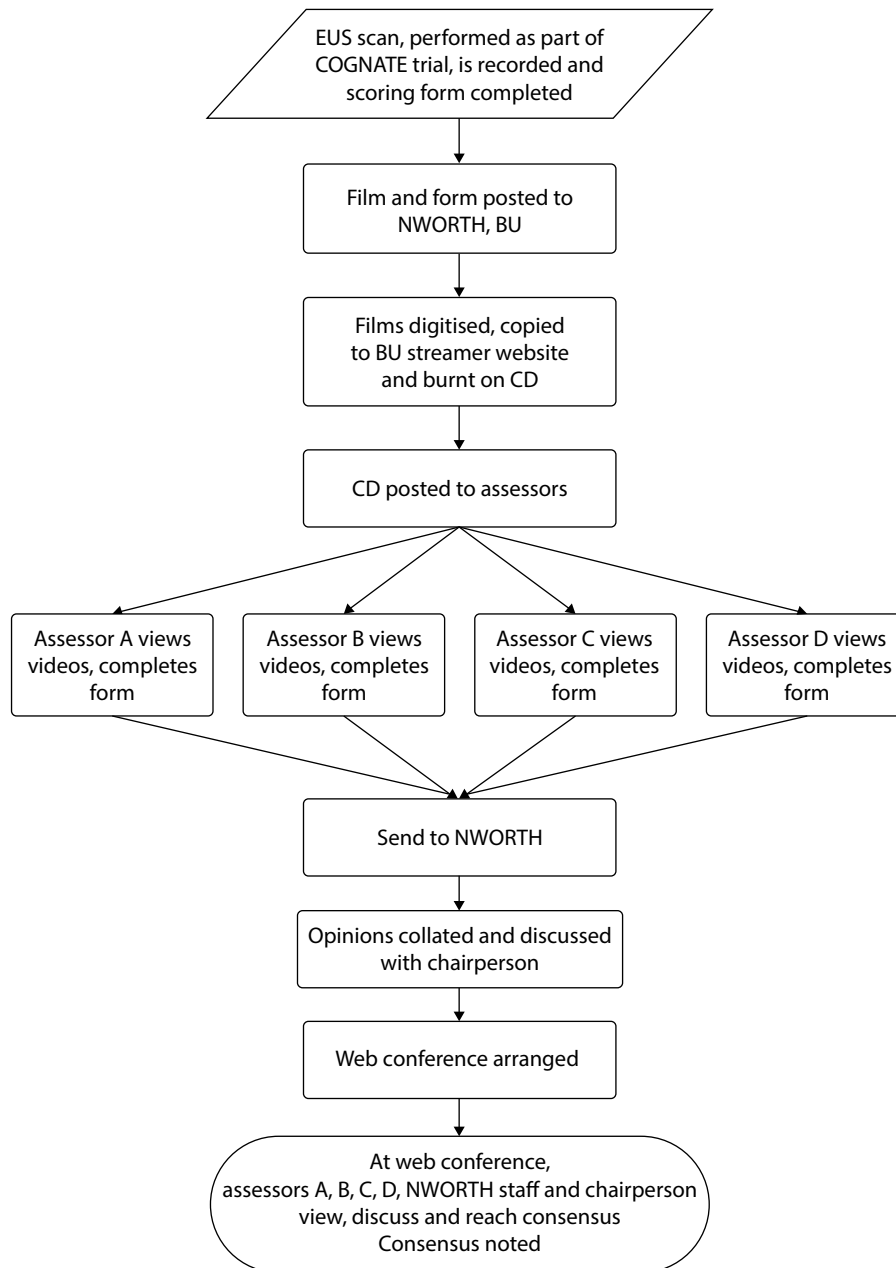
FIGURE 3 Process of quality assurance. BU, Bangor University.

Conscious of the danger that the panel, although blind to the source of the scans they were assessing, were favourably disposed to those scans, we recruited an experienced endosonographer from the Commonwealth as external assessor. We compared assessments between the original operator, panel members, the panel consensus and the external assessor, in principle by Cohen's kappa or weighted kappa.[130] Finally, for participants in the intervention group who had resection without pre-treatment, we compared EUS findings with pathological findings at operation.

Chapter 3 Results

Recruitment, participant flow and CONSORT

Recruitment

Identification and eligibility

Recruitment started in Aberdeen and Gloucester in September 2004, became fully established on 1 February 2005, and finished on 31 July 2008. Eight centres participated. The COGNATE team identified 1152 potential participants and recruited 223 (19.4%) to the trial (*Figure 4*, elaborated in *Appendix 4.1*).

The most common reason for exclusion was that patients were 'not medically fit for surgery' ($n = 383$). This reflects the physiological fitness of the patient, not the suitability of the tumour for surgery. Although only eight exclusions cited WHO status, many excluded as 'not medically fit for surgery' also had WHO status 3 or 4 so we combined these two categories. About half of those identified ($n = 567$) were initially adjudged eligible, given information about the trial and asked for consent. Of these 90 (15.9%) refused immediately; we sought no further details for them. The remaining 477 were discussed as potential participants at a multidisciplinary meeting (MDM) (see *Figure 1*), after which centres identified further exclusions. These included 24 further refusals, some of whom withdrew previous consent. Thus 114 people (9.9% of those identified; 20.0% of those potentially eligible) refused consent to the trial, leaving 453 potentially eligible.

Centres excluded a further 228 people after the MDM – a larger proportion than first predicted. We grouped reasons in free text into categories. Centres excluded 120 people because they regarded EUS as essential; for half of these, no reason was given. In contrast, only eight people apparently did not need an EUS. Another 36 patients had already received EUS or a booking for it. In all, we randomised 225 participants. Subsequent validity checks established that two were in error; we removed them from analysis, leaving 223.

Eligibility by centre

Differences between centres in numbers of patients identified, proportions recruited, and reasons for exclusion were highly significant. *Figures 5* and *6* summarise these differences after combining the four centres with fewest identified patients (full details are shown in *Appendix 4.2*). Although centres 20 and 22 identified more potential participants than other centres, they excluded many more than Aberdeen or Gloucester. Almost all of those with negative histology were from centres 29 and 30. 'Not fit for surgery' was least common in Gloucester, Aberdeen and centres 24 and 26. Gloucester had fewest exclusions for metastases. Almost all post-MDM refusals, but few pre-MDM refusals, came from Gloucester; this may reflect differing interpretations of consent. Unlike other centres, Aberdeen and Gloucester did not exclude people for 'mandatory' EUS scans. These findings led us to decide provisionally to combine all centres except Aberdeen and Gloucester into a single group for all analyses, and to test the validity of this plan through those analyses, notably by assessing how well data fitted this three-centre model.
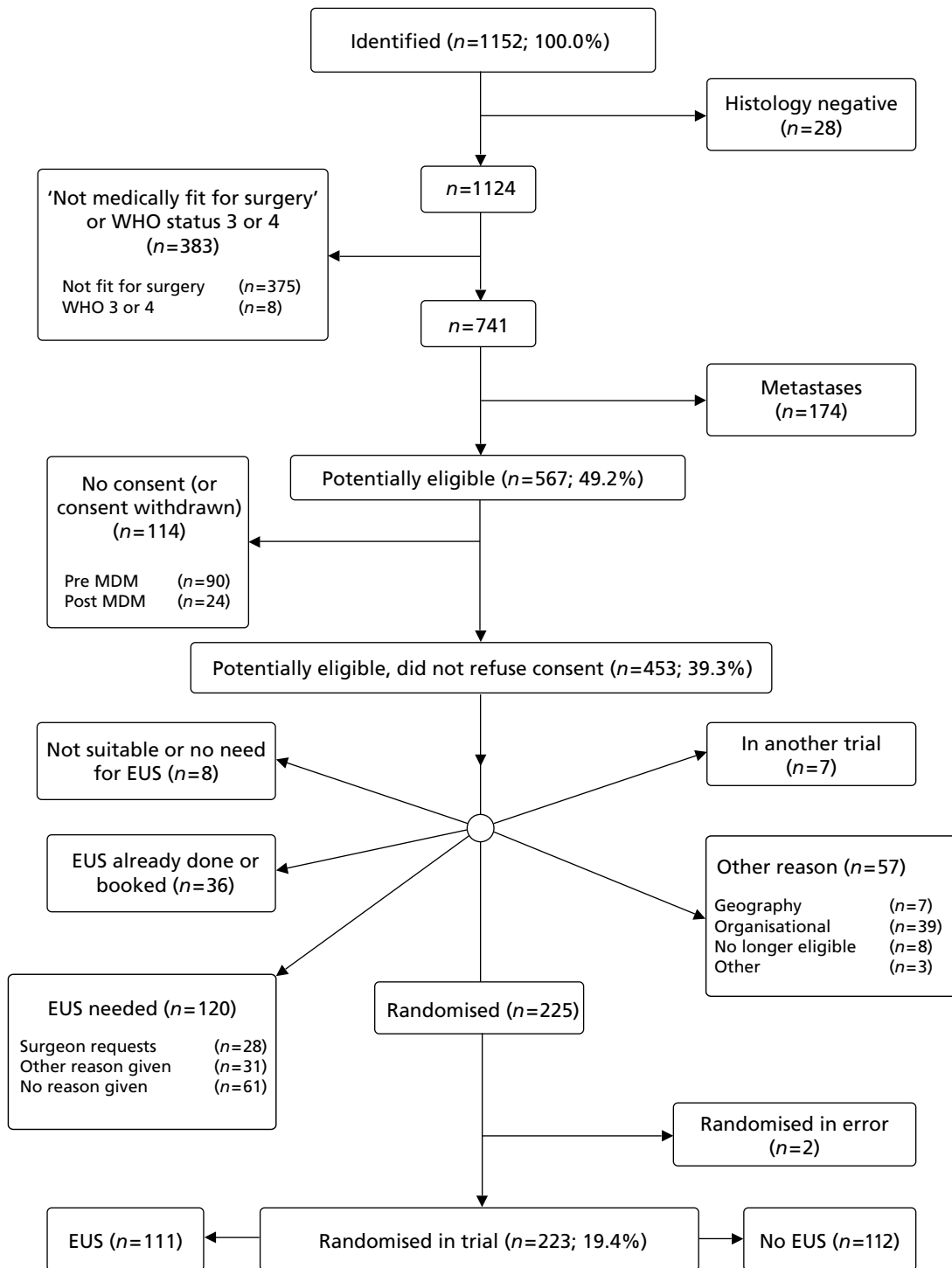
FIGURE 4 CONSORT flow chart to point of randomisation.
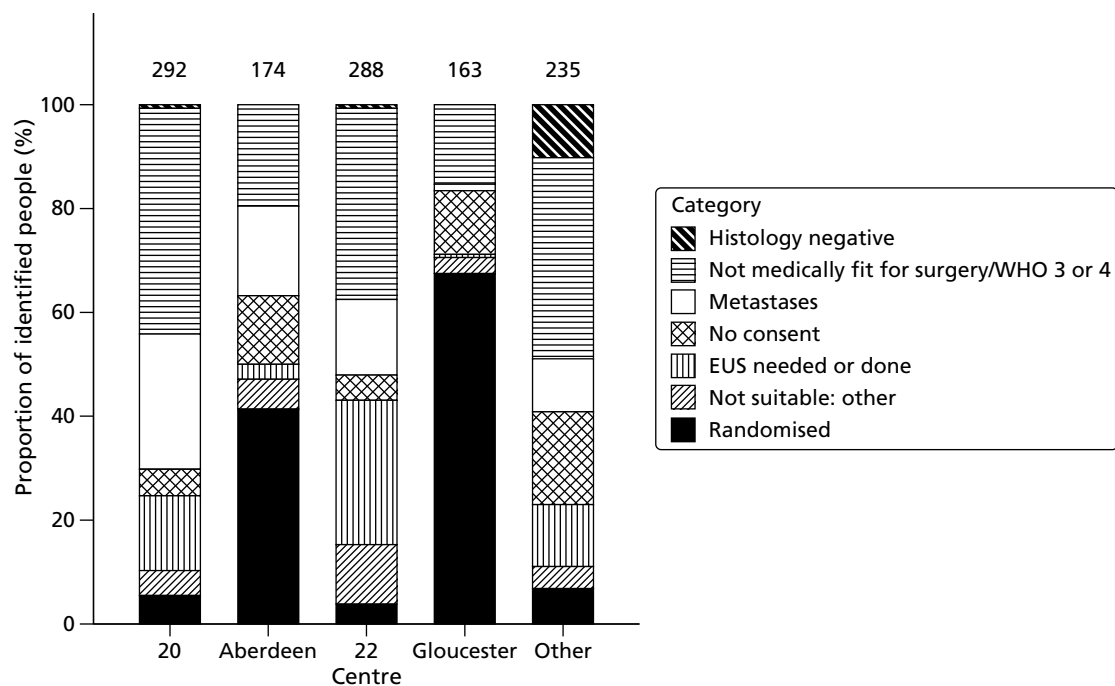
FIGURE 5 Patients identified by category within centre.
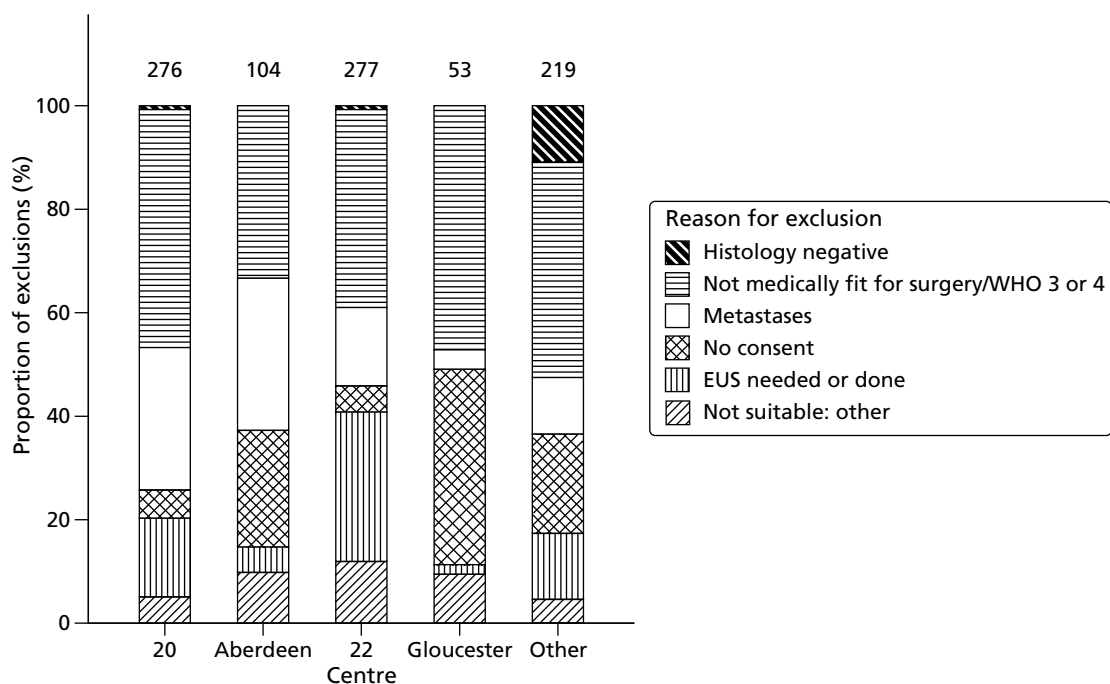
FIGURE 6 Patients excluded by category within centre.

Attrition, withdrawals and missing data
Figures 7 and *8*, and *Appendix 4.3*, show progress through the trial.

In summary, six participants (three in EUS group; three in non-EUS group) withdrew completely; 25 (13 in EUS group; 12 in non-EUS group) withdrew from questionnaires, but not from collection of hospital data. We lost baseline data for three participants when a laptop computer was stolen; baseline data were also unavailable for another participant. Although none of the four completed another questionnaire, none withdrew completely; all four later died. Another participant withdrew completely before any information could be input.

In accordance with the general principles of pragmatic trials (see *Table 7*), and the need to analyse by treatment allocated enshrined in the CONSORT (Consolidated Standards Of Reporting Trials) statement,[111] we sought to analyse as many participants as possible. Although 14 participants allocated to the EUS group did not receive a scan and three allocated to the non-EUS group did receive a scan, we analysed all 17 as allocated. Only a fastidious trial would label these two natural events as 'deviations from the protocol' (see *Table 7*). In contrast, we judge that these natural events strengthen the findings of our pragmatic trial as they represent normal clinical practice; to prevent them, even if possible and ethical, would transform COGNATE into a laboratory investigation of EUS that would be unable to yield the practical evaluation of the two policies – 'include EUS in management plan after MDT meeting' and 'do not include EUS in this plan' – that the NIHR HTA programme commissioned.
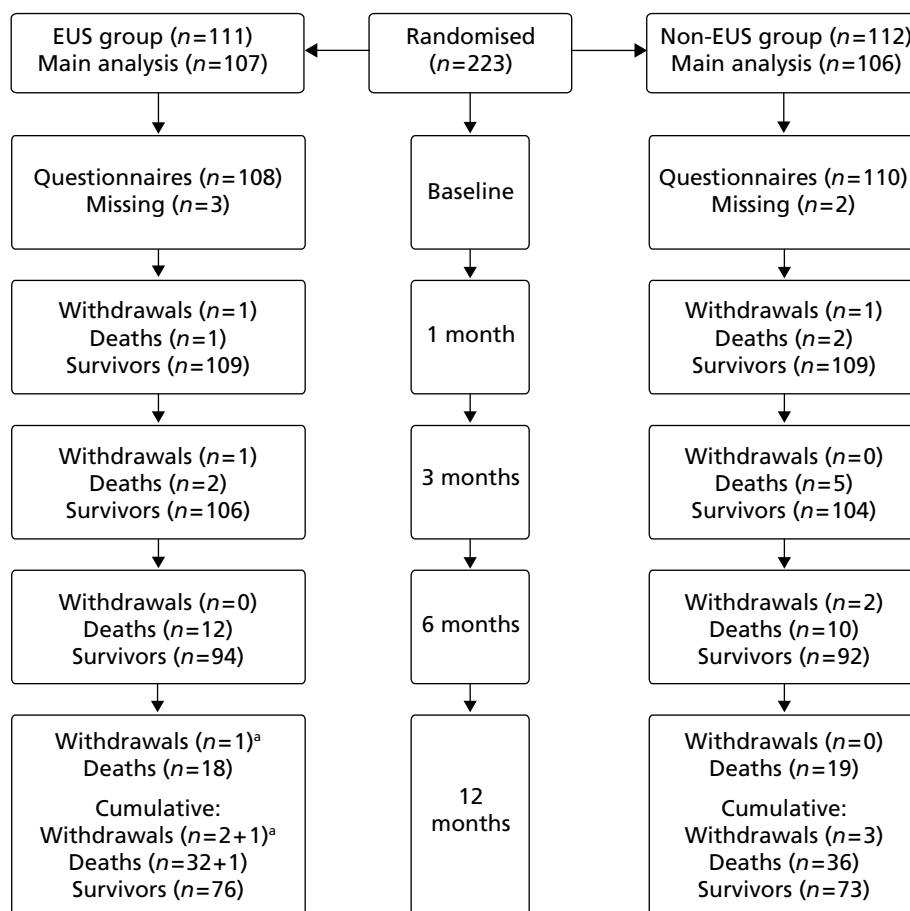
FIGURE 7 CONSORT flow chart (from randomisation to follow-up at 12 months). a, One participant died in the time period in which he or she withdrew.
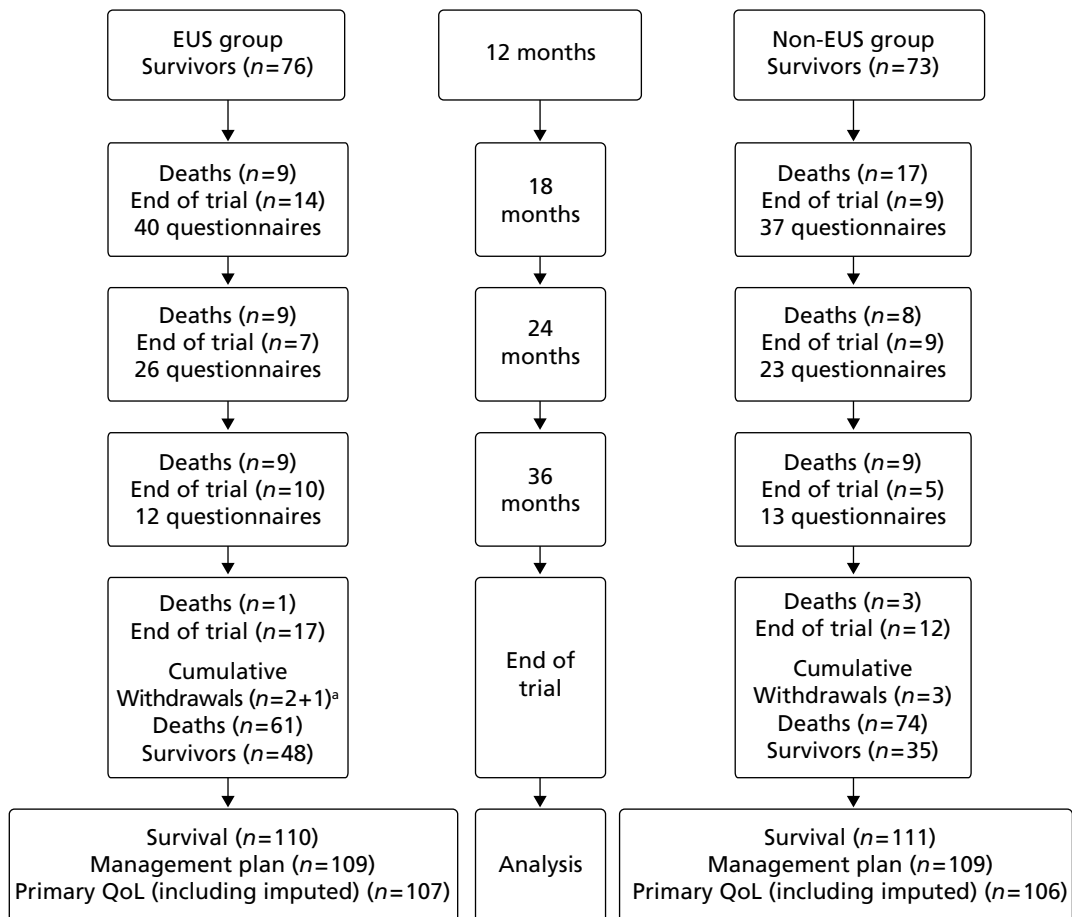
FIGURE 8 CONSORT flow chart (from follow-up at 12 months until end of trial). QoL, quality of life.

Furthermore, rather than remove participants who withdrew or provided incomplete information from analysis, we imputed their missing data whenever possible. In particular we know that 21 participants who withdrew died by the end of the trial. As EQ-5D is taken as zero after death, this information contributes to the primary outcome measure, QALYs, for these participants. However we excluded from primary analysis 10 participants with few or no useful data. Two withdrew completely before the 1-month interview and provided no information on outcome. The remaining eight provided no follow-up data, usually through non-attendance or withdrawal from interview, and either survived beyond the end of the trial or until they withdrew completely. Most of these participants also had little or no cost data. Thus 213 participants contributed to primary outcome analysis. To maximise power, however, we can compare survival and management plans (which do not require imputation of EQ-5D utilities from questionnaires or early deaths) on up to 221 participants.

Demographic and baseline characteristics

Table 9 compares baseline characteristics between intervention and control groups. As we stratified randomisation by centre and tumour site, these match very closely; the balance in other characteristics provides reassurance that the random allocation process was valid.

Over 75% of participants were male, and nearly half were aged over 65 years. The most common tumour site was oesophagus, the most common tumour type adenocarcinoma, and the most common tumour T stage was T3. Slightly more participants had nodal stage N1 than N0. As almost all were WHO status 1, we did not use this variable as a covariate. Similarly we reduced tumour site, T stage and type to binary

TABLE 9 Demographic and baseline data by allocated group for 213 participants in main analyses

Participant characteristic	Non-EUS group (*n* = 106): no. (%) unless stated	EUS group (*n* = 107): no. (%) unless stated	Whole sample (*n* = 213): no. (%) unless stated
Male	82 (77)	83 (78)	165 (77)
Age (years): mean (SD), minimum, maximum	64.3 (10.0) 26, 82	64.4 (9.4) 40, 83	64.4 (9.7) 26, 83
Age ≥65 years	50 (47)	50 (47)	100 (47)
WHO status 2	9 (8)	5 (5)	14 (7)
Tumour site			
Oesophagus	79 (75)	79 (74)	158 (74)
Junction	9 (8)	11 (10)	20 (9)
Gastric	18 (17)	17 (16)	35 (16)
Tumour T stage (pre-randomisation)			
Tis	0 (0)	1 (1)	1 (0)
T1	7 (7)	8 (8)	15 (7)
T2	22 (21)	26 (24)	48 (22)
T3	41 (39)	50 (47)	91 (43)
T4	5 (5)	3 (3)	8 (4)
Not recorded	31 (29)	19 (18)	50 (24)
Tumour N stage (pre-randomisation)			
N0	33 (31)	33 (31)	66 (31)
N1	39 (37)	48 (45)	87 (41)
N2	3 (3)	2 (2)	5 (2)
Not recorded or unknown	31 (29)	24 (22)	55 (26)
Tumour type			
Adenocarcinoma	86 (81)	87 (81)	173 (83)
Squamous	18 (17)	17 (16)	35 (16)
Other	2 (2)	3 (3)	5 (2)
Centre			
Aberdeen	34 (32)	36 (34)	70 (33)
Gloucester	58 (48)	51 (48)	102 (48)
Other centres	21 (20)	20 (19)	41 (19)

SD, standard deviation; Tis, tumour in situ.

variables for use as covariates. However pre-randomisation tumour stage was not recorded for one-quarter of participants. Very little other baseline information was missing among the 213 participants in the primary analysis.

Table 10 summarises the same baseline characteristics in the three centre groups, with the third group combining six separate centres. There were some significant differences between centres. Aberdeen participants were more likely to be aged under 65 years; those from Gloucester were the least likely to

TABLE 10 Demographic and baseline data by centre for 213 participants in main analyses

Participant characteristic	Aberdeen (*n* = 70): no. (%) unless stated	Gloucester (*n* = 102): no. (%) unless stated	Other centres (*n* = 41): no. (%) unless stated
Male	50 (71)	81 (79)	34 (83)
Age (years): mean (SD), minimum, maximum	62.9 (7.6) 47, 79	65.1 (10.6) 26, 83	64.9 (10.4) 45, 82
Age ≥65 years	25 (36)	54 (53)	21 (51)
WHO status 2	6 (9)	2 (2)	6 (15)
Tumour site			
Oesophagus	51 (73)	80 (78)	27 (66)
Junction	7 (10)	6 (6)	7 (17)
Gastric	12 (17)	16 (16)	7 (17)
Tumour T stage (pre-randomisation)			
Tis	0 (0)	0 (0)	1 (2)
T1	5 (7)	10 (10)	1 (2)
T2	19 (27)	9 (9)	19 (46)
T3	42 (60)	33 (32)	16 (39)
T4	4 (6)	3 (3)	1 (2)
Not recorded	0 (0)	47 (46)	3 (7)
Tumour N stage (pre-randomisation)			
N0	31 (44)	16 (16)	19 (46)
N1	38 (54)	30 (29)	19 (46)
N2	0 (0)	5 (5)	0 (0)
Not recorded or unknown	1 (1)	51 (50)	3 (7)
Tumour type			
Adenocarcinoma	60 (86)	79 (76)	34 (83)
Squamous	10 (14)	20 (20)	5 (12)
Other	0 (0)	3 (3)	2 (5)

SD, standard deviation; Tis, tumour in situ.

have WHO status 2. There was no record of pre-randomisation tumour stage for about half the Gloucester cases; among those with recorded T stage, Aberdeen and Gloucester had the highest proportion of T3 tumours. Despite the exclusion criteria, two biopsies from Aberdeen (both in the non-EUS group) were coded M1 and 11 more were equivocal.

Refinement of quality-of-life measures

Effective sample sizes
We immediately excluded one of the FACT Social items – 'I am satisfied with my sex life' – because the proportion of patients answering the question was low (40%, 43%, 37% at months 0, 1 and 3 respectively). After that exclusion and imputation (see *Chapter 2*, *Statistical methods*), the effective sample sizes for psychometric analyses were 220 at month 0, 173 at month 1 and 150 at month 3. These are

fewer than the numbers of survivors because some survivors did not attempt one or both questionnaires. Beyond month 3, the effective sample sizes were too small for psychometric analysis. Hence we based decisions about how to score FACT on the data for months 0, 1 and 3, and extrapolated our conclusions to the whole trial.

Principal components analysis of FACT-AC

Findings
The results of the principal components analysis of Additional Concerns for month 0 are in *Table 11* (with similar results for months 1 and 3 in *Appendices 5.1 and 5.2*). Together these analyses suggested seven dimensions. There was reasonably clear separation of eating problems and stomach problems. Other components were less clear. Within eating problems, there may be some redundancy of items. Given this redundancy, we had three options for computing an individual's Additional Concerns score: first, we could use all the items for all the participants (full combined scale); secondly we could eliminate clearly redundant items and use the remainder for all participants (abridged combined scale); or thirdly we could use the original oesophageal items for the oesophageal patients and the original gastric items for the gastric patients (separate scales). To throw light on this choice, we experimented with abridging the scale by removing clearly redundant items (marked by superscript 'c' in *Table 11*) leaving 21 items. When we repeated the principal components analysis, essentially the same seven dimensions emerged. However at month 0 the full combined scale correlated 0.96 with the separate scales but only 0.80 with the abridged combined scale; and the abridged combined scale correlated only 0.72 with the separate scales. The correlation patterns at months 1 and 3 were similar.

Implication for COGNATE: how to score FACT-AC?
We expected FACT-AC to show many dimensions because it tries to summarise not one but many diverse concerns, each of which might affect quality of life. Arguably, the items are better viewed as formative than as reflective indicators.[157] If so, we can legitimately score it as one or two additive rating scales. The redundancy of items is in part the result of our combining two similar scales, and in part inherent in the original scales. That is why we have three choices – full or abridged combined scale, or original scales. As COGNATE combines oesophageal and gastric patients, a combined scale is more appropriate.

However combining the scales alters the original weightings of the various concerns. For example oesophageal patients completing the combined scale will answer relatively fewer questions about swallowing than if they had answered the oesophageal scale. Abridging the combined scale further alters the weightings of the various concerns. This is why the full combined scale correlated more highly with the original scales than the abridged combined scale did. In these circumstances, we decided it was prudent to use the full combined scale for outcome analysis in the COGNATE trial.

Confirmatory factor analysis of FACT-G

Findings
The results of the confirmatory factor analyses for month 0 are in *Table 12* (with similar results for months 1 and 3 in *Appendices 5.3 and 5.4*). Generally they confirmed the four-factor structure. However we made modifications of two kinds. First, we allowed some measurement errors to covary within subscales, although not between subscales. For example, within the Social subscale at each time point, we allowed the items 'My family has accepted my illness' and 'I am satisfied with family communication about my illness' to correlate positively. Within the Functional subscale at each time point, we allowed the items 'I am able to work (include work at home)' and 'My work (include work at home) is fulfilling' to correlate positively. Second, we permitted some items to load on more than one factor. For example, we gave the supposedly Functional item 'I have accepted my illness' more flexibility: at month 0 it loaded 0.30 on Functional but also –0.33 on Emotional; at month 1 it loaded 0.48 on Functional but also 0.28 on Social; and at month 3 it loaded 0.27 on Functional but also –0.32 on Emotional. We also gave the supposedly Physical item 'I am bothered by the side effects of treatment' more flexibility: at month 0 it loaded 0.06

TABLE 11 Principal components analysis of FACT Specific Concerns at month 0[a]

Item	Components						
	1[b]	2[b]	3[b]	4[b]	5[b]	6[b]	7[b]
Item–component loadings							
Ga12.0 I have trouble swallowing food	0.89	0.05	0.09	0.00	0.03	−0.01	0.06
E1.0 I have difficulty swallowing solid foods[c]	0.87	0.04	0.06	0.04	0.01	−0.06	0.03
HN1.0 I am able to eat the foods that I like	−0.76	−0.07	−0.06	−0.31	−0.14	0.05	0.06
Ga9.0 I avoid going out to eat because of my illness[c]	0.75	0.04	0.08	0.12	0.20	0.00	−0.03
HN7.0 I can swallow naturally and easily[c]	−0.72	0.05	−0.06	−0.09	0.04	−0.18	0.13
Ga6.0 I have discomfort or pain when I eat	0.72	0.27	−0.06	0.13	−0.04	−0.09	0.06
Ga4.0 I am bothered by a change in my eating habits	0.66	0.11	0.03	0.11	0.26	0.11	0.08
E6.0 I am able to enjoy meals with family or friends	−0.66	−0.10	−0.02	−0.25	−0.22	−0.14	0.15
HN5.0 I am able to eat as much food as I want[c]	−0.65	0.01	−0.10	−0.43	−0.12	−0.03	0.13
E 2.0 I have difficulty swallowing soft or mashed foods[c]	0.63	−0.08	0.01	0.08	0.05	0.34	0.37
Ga10.0 My digestive problems interfere with my usual activities[c]	0.61	0.21	0.08	0.17	0.42	0.03	0.02
E5.0 I choke when I swallow	0.56	0.22	0.19	−0.23	−0.17	−0.04	−0.28
E4.0 I have pain in my chest when I swallow	0.51	0.44	0.00	−0.14	−0.21	−0.12	0.07
C2.0 I am losing weight	0.47	−0.08	0.23	0.46	0.20	−0.07	0.17
ACT11.0 I have pain in my stomach area[c]	0.09	0.79	0.16	0.02	0.16	0.07	0.02
Hep8.0 I have discomfort or pain in my stomach area	0.16	0.77	0.13	−0.03	0.26	0.08	0.13
C1.0 I have swelling or cramps in my stomach area[c]	−0.01	0.60	0.18	−0.07	0.07	0.13	−0.18
Ga5.0 I have a feeling of fullness or heaviness in my stomach area[c]	0.01	0.57	0.00	0.45	0.10	0.11	0.03
Ga14.0 I am bothered by gas (flatulence)	0.24	0.43	0.09	0.11	−0.12	−0.22	−0.22
E7.0 I wake at night because of coughing	0.05	0.02	0.72	−0.09	0.11	0.07	−0.02
HI12.0 I feel weak all over	0.08	0.11	0.71	0.34	0.07	0.07	−0.04
An2.0 I feel tired	0.16	0.20	0.64	0.25	0.18	−0.12	−0.12
HN2.0 My mouth is dry	0.00	0.24	0.57	0.10	−0.07	0.19	0.10
C6.0 I have a good appetite[c]	−0.37	0.02	−0.17	−0.66	−0.06	−0.14	0.11
Ga1.0 I have a loss of appetite	0.34	0.07	0.23	0.62	0.02	−0.09	−0.04
Leu4.0 Because of my illness, I have difficulty planning for the future[c]	0.22	0.17	0.22	0.06	0.69	−0.05	−0.07
Ga7.0 I worry about having stomach problems[c]	0.18	0.33	0.03	0.19	0.58	0.31	0.18
HN10.0 I am able to communicate with others	0.00	−0.06	−0.05	−0.06	−0.09	−0.70	0.19
E3.0 I have difficulty swallowing liquids	0.46	0.15	0.24	−0.06	−0.09	0.50	0.31
HN3.0 I have trouble breathing	−0.06	0.31	0.15	0.32	−0.26	0.41	−0.24
HN4.0 My voice has its usual quality and strength	−0.15	−0.10	−0.21	0.22	−0.31	−0.37	0.20
C5.0 I have diarrhoea	−0.06	−0.07	−0.07	−0.02	0.09	0.19	−0.71
Ga2.0 I am bothered by reflux or heartburn	0.14	0.19	0.16	0.19	−0.13	0.04	−0.43

a *n* = 220.

b These seven components explain 58% of the variance.

c Items omitted from abridged combined scale.

TABLE 12 Confirmatory factor analysis of FACT-G at month 0[a,b]

Item	Factors			
	Physical	Social	Emotional	Functional
Item–factor loadings				
GP1. I have a lack of energy	0.63			
GP2. I have nausea	0.52			
GP3. Because of my physical condition, I have trouble meeting the needs of my family	0.62			
GP4. I have pain	0.60			
GP5. I am bothered by side effects of treatment	0.06		0.26	
GP6. I feel ill	0.80			
GP7. I am forced to spend time in bed	0.69			
GS1. I feel close to my friends[c]		0.32		
GS2. I get emotional support from my family		0.79		
GS3. I get support from my friends[c]		0.47		
GS4. My family has accepted my illness[c]		0.35		
GS5. I am satisfied with family communication about my illness[c]		0.49		
GS6. I feel close to my partner (or the person who is my main support)		0.63		
GE1. I feel sad			0.70	
GE2. I am satisfied with how I am coping with my illness			−0.42	
GE3. I am losing hope in the fight against my illness			0.39	
GE4. I feel nervous			0.77	
GE5. I worry about dying			0.71	
GE6. I worry that my condition will get worse			0.67	
GF1. I am able to work (include work at home)[c]				0.56
GF2. My work (include work at home) is fulfilling[c]				0.61
GF3. I am able to enjoy life				0.83
GF4. I have accepted my illness			−0.33	0.30
GF5. I am sleeping well				0.47
GF6. I am enjoying the things I usually do for fun				0.83
GF7. I am content with the quality of my life right now				0.80
Factor–factor correlations				
Physical	–			
Social	−0.06	–		
Emotional	0.14	−0.04	–	
Functional	−0.70	0.28	−0.34	–

a $n = 220$.

b $\chi^2(288) = 502.09$, $p < 0.01$; CFI = 0.90; RMSEA = 0.06, SRMR = 0.08; p (close fit) = 0.06.

c Permitted correlations between measurement errors: GS1 positively with GS3; GS4 positively with GS5; GF1 positively with GF2.

on Physical but also 0.26 on Emotional; at month 1 it loaded 0.57 on Physical but also 0.21 on Emotional; and at month 3 it loaded 0.53 on Physical but also 0.29 on Emotional. No other item showed a tendency to cross-load at more than one single time point. There was a consistently high correlation between Physical and Functional Well-Being factors: −0.70 at month 0, −0.65 at month 1 and −0.78 at month 3.

Implication for the trial: how to score FACT-G?

The high proportion of missing data shows that the item 'I am satisfied with my sex life' is particularly sensitive. Indeed respondents are already asked if they wish to skip this question. The cross-loadings in the confirmatory factor analysis suggest that the item 'I have accepted my illness' does not adequately reflect functional well-being. Although the item 'I am bothered by the side effects of treatment' did not adequately reflect physical well-being on entry to the trial, it later did so. One simple explanation would be that, with the passage of time, patients had undergone more treatment. The correlated measurement errors suggested slight heterogeneity within subscales. As this heterogeneity seems to reflect the specific contexts to which items referred, notably family or work, within otherwise unitary constructs, it was less of a concern. We therefore decided to drop two items – 'I am satisfied with my sex life' and 'I have accepted my illness'. That apart, there is no reason to depart from the established four-dimensional structure. When combining FACT subscales, we would continue to give equal weight to each subscale.

Structural equation modelling of relationships between measures

Findings

The internal consistencies and correlations between scales for month 0 are in *Table 13* (with similar results for months 1 and 3 in *Appendices 5.5* and *5.6*): the internal consistencies of all multi-item scales were at least 0.70 at each time point; and EQ-5D and EQ-VAS correlated 0.40, 0.55 and 0.63 at months 0, 1 and 3 respectively. We ran separate structural equation models for EQ-5D and EQ-VAS. Those for EQ-5D in month 0 are in *Figure 9* (with similar results for months 1 and 3 in *Appendices 5.7* and *5.8*): EQ-5D was predicted by Physical Well-Being at each time point, and also by Functional Well-Being at months 1 and 3. Additional Concerns predicted Physical, Functional, and Emotional Well-Being at all three time points, and also Social Well-Being at month 1. There was no evidence for a direct effect of Additional Concerns on EQ-5D, that is an effect not mediated by Physical and Functional Well-Being. The models for EQ-VAS in month 0 are in *Figure 10* (with similar results for months 1 and 3 in *Appendices 5.9* and *5.10*). They show the same structure as for EQ-5D.

TABLE 13 Internal consistency and intercorrelations of scales at month 0[a]

Scale	Internal consistency	Correlations					
		1	2	3	4	5	6
1. FACT Specific Concerns	0.91	–					
2. FACT Physical	0.75	−0.50**	–				
3. FACT Social	0.73	−0.11	0.11	–			
4. FACT Emotional	0.78	−0.26**	0.19**	0.12	–		
5. FACT Functional	0.84	−0.50**	0.61**	0.26**	0.29**	–	
6. EQ-5D	–	−0.28**	0.53**	0.06	0.25**	0.44**	–
7. EQ-VAS	–	−0.45**	0.62**	0.15*	0.04	0.50**	0.40**

*p<0.05; **p<0.01.
a   n = 220.
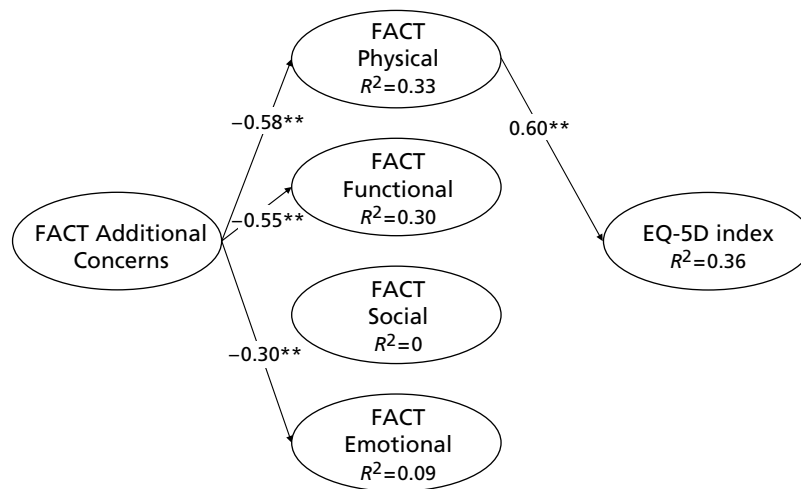
FIGURE 9 FACT predicting EQ-5D at month 0. [$\chi^2_4 = 6.52$; $p = 0.16$; SRMR = 0.04; CFI = 0.98; RMSEA = 0.07; p (close fit) = 0.21]. **$p < 0.01$.
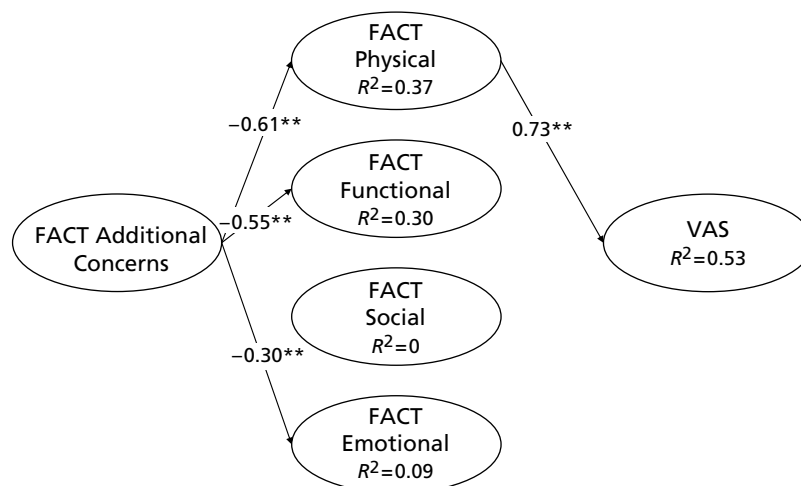
FIGURE 10 FACT predicting EQ-VAS at month 0. [$\chi^2_5 = 11.03$; $p = 0.05$; SRMR = 0.04; CFI = 0.98; RMSEA = 0.07; p (close fit) = 0.21]. **$p < 0.01$.

Implication for the trial: what FACT summary measure to use?

Given that only FACT Physical and Functional Well-Being predicted EQ-5D, it seems that EQ-5D did not capture individuals' social or emotional well-being. This could be because the EQ-5D items do not adequately cover these domains, or it could be because the public (who provide the weightings underpinning the EQ-5D score) do not think these are important domains. The EQ-VAS performed like the EQ-5D, perhaps suggesting that the patients themselves give little weight to social and emotional well-being when assessing their general quality of life. However patients' interpretation of the EQ-VAS could be biased by the content of the immediately preceding EQ-5D.

We recall that we included FACT in our outcomes mainly as an alternative to the EQ-5D as a means of adjusting survival. Therefore it would be counterproductive to use the FACT TOI, which includes the subscales that predict EQ-5D (Physical and Functional) and excludes those that do not (Social and Emotional). That would duplicate, not complement, what we might learn from the EQ-5D. Given that Additional Concerns predicted Physical, Functional, and Emotional (and, at one time point, Social) Well-Being, it would also be counterproductive to use FACT Total, which merely adds FACT-AC to FACT-G. That would in effect sum cause and effect. Hence, in addition to adjusting survival by EQ-5D, we decided to give ourselves the best chance of new insights by adjusting survival both by FACT-G and by FACT-AC.

Effectiveness

Survival

Proportion surviving

Five participants withdrew completely so we do not know whether they were alive at the end. However three who remained in the trial for longer than the first participant who died contributed to survival analysis. Of the remaining 109 intervention patients 48 (44%) survived until the end of the trial, compared with 35 (32%) of the remaining 109 control patients [relative risk (RR) = 0.84; 95% CI from 0.67 to 1.02; $p = 0.070$]. Also 76 (70%) intervention patients survived until 12 months, compared with 73 (67%) control patients (RR = 0.92; 95% CI from 0.62 to 1.36; $p = 0.66$).

Survival curves: univariate comparisons

Figure 11 shows the observed survival curves for the two groups. These take account of withdrawals and different lengths of time in the trial (between 1 and some 4 years) to provide unbiased estimates of the proportion of the original sample still alive at each time. Because of censoring, the lower portion of the curve suffers more random variation. For example a single death represents a change in survival of about 0.01 in the first year of the trial and about 0.02 at 3 years.

The intervention survival curve lies almost completely above the control curve. Survival is similar in the two groups up to about 1 year, after which the curves diverge. *Figures 12–16* subdivide survival curves by other baseline characteristics.

Survival was significantly better in early tumour stages (univariate log-rank test; $p = 0.017$), and raised but not significantly better for non-squamous tumours and participants under 65 ($p = 0.11$ and 0.14 respectively). However centre, gender and tumour site do not contribute to survival in isolation ($p = 0.48$, 0.51 and 0.37 respectively). As expected survival was strongly related to management plan before randomisation and hence before EUS ($p = 0.005$): those planned for surgery or EMR survived longest, whereas those scheduled for chemotherapy or radiotherapy fared worst. Some survival curves are more
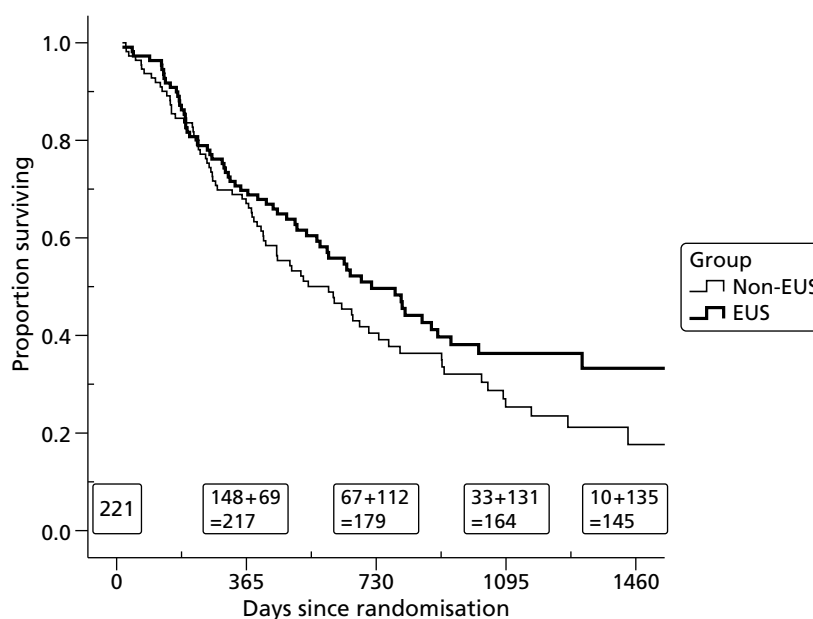
FIGURE 11 Survival curves by allocated treatment group. Note: rectangular boxes above the *x*-axis record the number of participants available to estimate survival curve, subdivided into participants still alive + cumulative deaths; thus these boxes exclude participants censored when the trial ended or they withdrew completely.
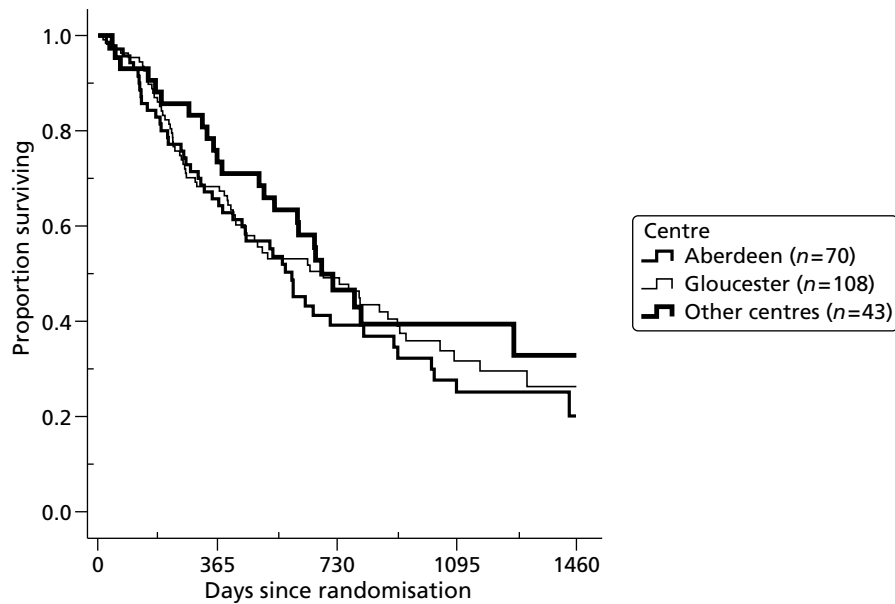
FIGURE 12 Survival curves by centre.
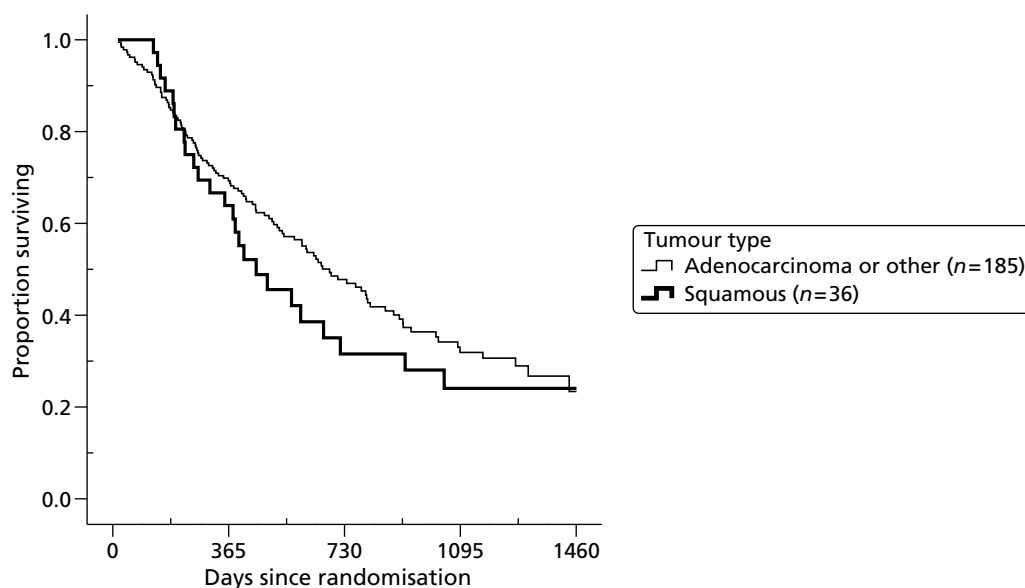
FIGURE 13 Survival curves by tumour type.

open to random variation, for example those comparing participants from 'other' centres ($n = 43$) or with gastric or junctional tumours ($n = 60$). Unfortunately, more than 50 baseline tumour stages were absent from the patients' notes from which we extracted our electronic data; their survival pattern was similar to that of the T3–T4 group.

Table 14 compares estimated mean and median survival times in intervention and control groups. Like survival curves, these estimates take account of varying follow-up times among survivors, so differ from the average length of survival. As there were no deaths after 4 years, we truncated the survival times of the few who were known to have survived beyond that time to 1461 days for consistency with the cost-effectiveness analysis. Estimated mean values still depend on the length of the trial, so medians are
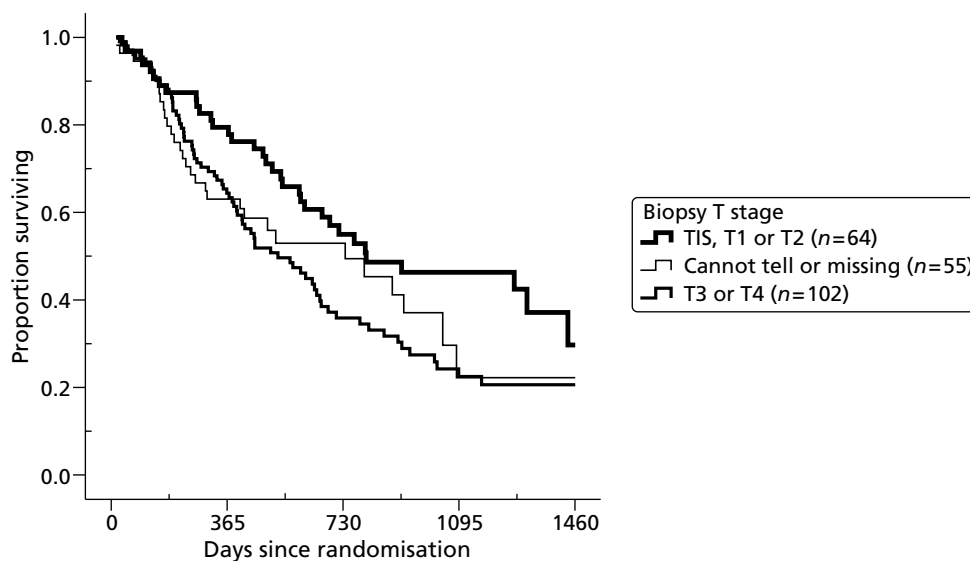
FIGURE 14 Survival curves by tumour stage (T).
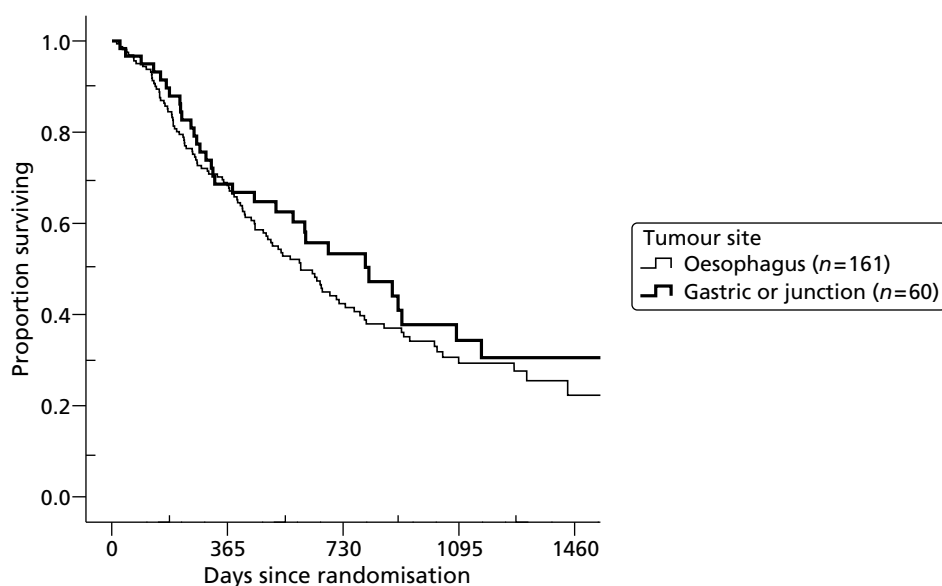
FIGURE 15 Survival curves by tumour site.

more robust. The median survival times were 20 months for the control group, and 24 months for the intervention group.

Table 15 gives a similar comparison between centres. The median survival times were 20 months in Aberdeen and 23 months in both Gloucester and the other centres.

Survival models with multiple predictors
Survival analysis takes account of censoring of observations at the time of withdrawal from the trial or other loss to follow-up. Our analysis therefore included all 221 people who had useful survival data, namely those who survived longer than the first participant who died.
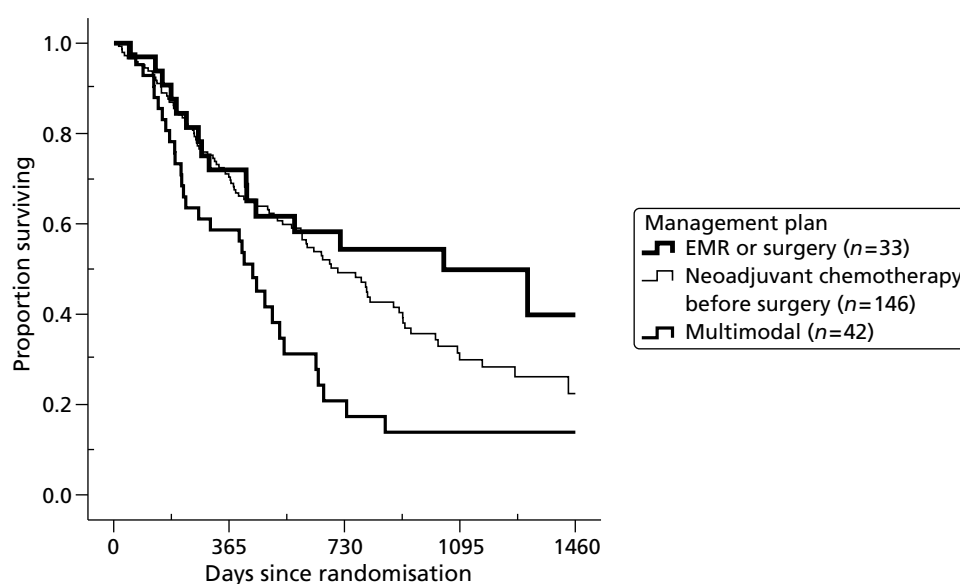
FIGURE 16 Survival curves by management plan.

TABLE 14 Estimated mean and median survival time by allocated group

Group	Mean survival in days (truncated at 1461 days) Mean (95% CI)	Median survival in days Median (95% CI)
Non-EUS	700 (601 to 799)	596 (415 to 777)
EUS	813 (707 to 919)	717 (533 to 901)
Both	755 (683 to 828)	647 (520 to 774)

TABLE 15 Estimated mean and median survival time by centre

Centre	Mean survival in days (truncated at 1461 days) Mean (95% CI)	Median survival in days Median (95% CI)
Aberdeen	709 (583 to 834)	592 (484 to 700)
Gloucester	761 (655 to 867)	688 (396 to 980)
Other centres	836 (672 to 1001)	683 (530 to 836)
All centres	755 (682 to 828)	647 (520 to 774)

Cox regression uses a logarithmic model to estimate the hazard ratio, which is the ratio of the instantaneous chance of dying in intervention and control groups. We specifically chose a proportional hazards model, in which this ratio is the same at all time points. If that hazard ratio were 1.0, there would be no difference between groups: at any time intervention and control participants would have an equal chance of dying. If we were to use only the allocated treatment to predict risk, the estimated hazard ratio would be 0.76, with a 95% confidence interval from 0.54 to 1.06 (see *Appendix 6.1*). At any time after randomisation, the estimated risk of death for an intervention participant would be only three-quarters of that for a control participant. As the confidence interval for that hazard ratio includes 1.0, however, we could not be 95% certain that the risk of death is lower with EUS than without.

Consistent with our agreed analysis plan (see *Chapter 2*, *Statistical methods*), we extended our model to allow for covariates. As potential covariates we considered centre, age, gender, tumour type and site, biopsy tumour (T) stage, baseline EQ-5D and management plan before randomisation. To keep the model parsimonious, we dichotomised pre-randomisation T stage into T2 or less versus T3 or more [including not recorded (NR)], and management plan into multimodal versus not. Although *Figure 12* suggests that survival may be similar across centres, the known differences between centres (see *Figures 5* and *6*) led us to split them into two contrasts – Gloucester and Aberdeen versus the other centres, and Gloucester versus Aberdeen. Adding T stage, multimodal plan and centre thus defined to the model improved the EUS hazard ratio, which became significant (*Table 16*, plus see *Appendix 6.2*).

Two examples may help interpretation. First, if we compare two participants who differ only in whether allocated to EUS, the chance of the intervention participant dying at any time is 71% of the chance of the corresponding control participant with the same centre, initial T stage, management plan and baseline quality of life. Second, if we compare a control participant with a multimodal plan (and personal characteristics associated with that) with an intervention participant with a different plan (and personal characteristics, although in the same centre with the same initial T stage and baseline quality of life) then our model tells us that the latter's relative chance of death is 0.706 divided by 1.415, namely one half.

Building on the model in *Table 16*, *Figure 17* shows the estimated population survival curves when all covariates are at their mean. The pattern is similar, but not identical, to the corresponding empirical survival curves in *Figure 12*. This similarity confirms that our data are consistent with the proportional hazards assumption in the Cox regression model.

We considered whether or not to add interactions with covariates to this model. None of the main covariates in that model had a significant interaction with the effect of EUS on survival. However baseline EQ-5D, although not significant in its own right, was significantly associated with the effect of EUS, yielding the hazard ratios shown in *Table 17*. This interaction improves the fit of the model, and the main effect of EUS remains significant.

To help interpretation, we considered three participants with the same centre, pre-randomisation T stage and management plan, but different baseline EQ-5D scores – low (0.6), average (0.8) and high (1.0). Hence most of the hazard ratios in *Table 17* do not contribute to relative chances. By analogy with the first example above, the second participant has an instantaneous risk of dying after EUS that is 70% of the risk without such a scan. In contrast the first participant has an EQ-5D score of 0.6, that is -0.2 below the mean EQ-5D. Thus the interaction term multiplies the hazard ratio by $(13.18)^{-0.2} = 0.597$, thus changing it to 42%, namely 0.597×0.704. Also in contrast the third participant has an EQ-5D score of 1.0, that is 0.2 $(= 1.0 - 0.8)$ above the mean EQ-5D. Thus the interaction term multiplies the hazard ratio by $(13.18)^{0.2} = 1.675$, thus changing it to 118%, namely 1.675×0.704. Lest an increase of 18% in mortality

TABLE 16 Survival by allocated group: Cox regression estimates without interaction

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS	**0.706**	**0.501 to 0.996**	**0.048**
Other centres vs (Aberdeen or Gloucester)	0.929	0.723 to 1.194	0.567
Gloucester vs Aberdeen	0.858	0.707 to 1.041	0.120
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.353**	**1.097 to 1.670**	**0.005**
Multimodal plan vs not	**1.415**	**1.147 to 1.746**	**0.001**
Baseline EQ-5D	0.716	0.277 to 1.849	0.490

Tis, tumour in situ.
Rows in bold type contribute significantly to the model.
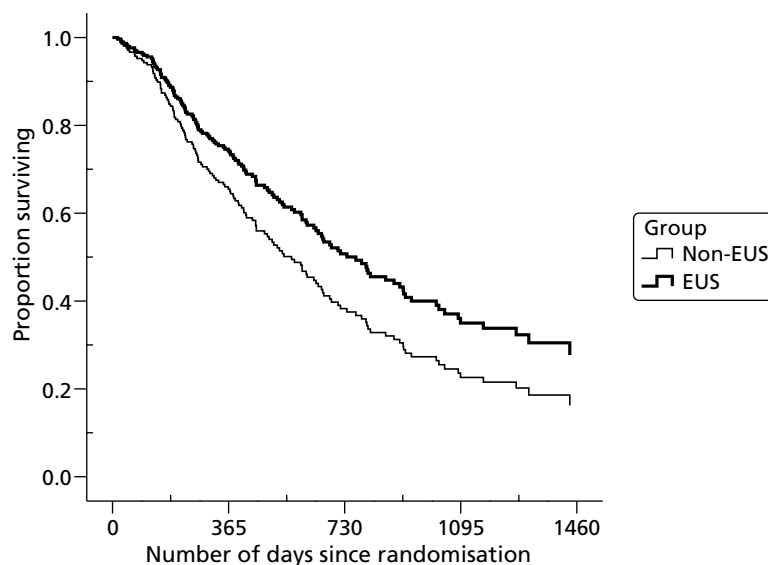
FIGURE 17 Survival curves estimated from Cox model without interaction by group.

TABLE 17 Survival by allocated group: Cox regression estimates with interaction

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS	**0.704**	**0.499 to 0.993**	**0.046**
Other centres vs (Aberdeen or Gloucester)	0.938	0.741 to 1.188	0.596
Gloucester vs Aberdeen	0.865	0.713 to 1.051	0.144
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.286**	**1.039 to 1.591**	**0.021**
Multimodal plan vs not	**1.443**	**1.168 to 1.782**	**0.001**
Baseline EQ-5D	0.518	0.189 to 1.424	0.203
Interaction: EUS and EQ-5D	**13.18**	**1.672 to 107.0**	**0.016**

Tis, tumour in situ.
Rows in bold type contribute significantly to the model.

risk seem alarming, we recall that this estimate is subject to enough random error that we cannot conclude that EUS increases the risk of dying even for the subgroup with baseline utility of 1. In short survival averaged over all participants is better in the intervention group; and lower baseline EQ-5D scores give greater advantage to the intervention.

Quality of life

Table 18 shows no significant differences in baseline quality-of-life scores between groups. However there is evidence of systematic differences between centres in quality of life at baseline; in particular Gloucester had the worst baseline (see *Appendix 6.3*). However the difference is only significant for FACT-G, FACT Social and EQ-VAS. We judge that the different proportions excluded by the three 'centres' and the resulting differences in case mix explain these baseline differences.

In comparing EQ-5D utility scores at baseline and follow-up between groups, we initially used only participants with interview data (face to face, telephone or postal) at those times (see *Appendix 6.4*). We imputed the few individual EQ-5D items omitted before calculating their utility scores (see *Chapter 2*, *Imputation and missing values*, *Imputing quality of life: phase 1 – psychometric and effectiveness*

TABLE 18 Baseline quality of life by allocated group for 213 participants in main analysis

Scale	Non-EUS group (*n* = 107) Mean (SD) minimum, maximum	EUS group (*n* = 106) Mean (SD) minimum, maximum	Whole sample (*n* = 213)[a] Mean (SD) minimum, maximum
FACT (range 0–4, 4 best)			
Physical	3.39 (0.53) 1.6, 4.0	3.38 (0.64) 0.6, 4.0	3.38 (0.58) 0.6, 4.0
Social	3.41 (0.71) 0.2, 4.0	3.44 (0.71) 1.7, 4.0	3.42 (0.63) 0.2, 4.0
Emotional	3.00 (0.73) 1.0, 4.0	2.94 (0.79) 0.0, 4.0	2.97 (0.76) 0.0, 4.0
Functional	2.88 (0.96) 0.2, 4.0	2.85 (0.96) 0.2, 4.0	2.86 (0.96) 0.2, 4.0
General (mean of four scores)	3.17 (0.49) 1.4, 4.0	3.15 (0.51) 1.5, 4.0	3.16 (0.50) 1.4, 4.0
FACT-AC (range 0–4, 0 best)	1.19 (0.65) 0.0, 2.6	1.02 (0.68) 0.0, 3.2	1.10 (0.67) 0.0, 3.2
EQ-5D utility score (range –0.6 to 1.0, 1.0 best)	0.80 (0.16) 0.1, 1.0	0.81 (0.20) –0.1, 1.0	0.80 (0.18) –0.1, 1.0
EQ-VAS (range 0–100, 100 best)	70.1 (18.8) 20, 100	74.4 (18.0) 20, 100	72.3 (18.5) 20, 100

SD, standard deviation.

a Including four participants whose baseline values were imputed from other information.

analyses). In general, respondents' EQ-5D scores fell in the first 3 months and then remained stable until 24 months, although the number of respondents was decreasing. There were no consistent differences between intervention and control groups.

Other quality-of-life scores had changed less at 12 months (see *Appendix 6.5*), with FACT and EQ-VAS means similar to those at baseline. Mean FACT subscale scores varied from about 2.5 for FACT Functional to about 3.5 for FACT Social. The EQ-VAS means were at about three-quarters of their maximum value. Adding imputed scores for surviving non-responders increases the number of participants at 12 months from 112 to 143 (see *Appendix 6.6*). Mean scores are similar to, but marginally worse than, the corresponding means for responders only (see *Appendix 6.5*). This suggests that non-responders have slightly worse health.

Throughout follow-up more control participants than intervention participants had died (*Figure 11* and see *Appendix 6.1*). Hence comparisons of survivors are likely to be biased against the intervention group. As EQ-5D explicitly assigns a utility of zero to death, *Table 19* uses this for known non-survivors at all times between their deaths and the end of the trial. Combining these zeros with the non-zero scores of survivors gives an unbiased comparison between groups. *Table 19* has two rows for each of the 18-, 24- and 36-month follow-ups. The first uses only participants whose survival status (i.e. whether dead or alive) is known, but imputes quality of life if an interview has been missed. To avoid bias, it excludes deaths recruited near the end of the trial who would not have reached this time point if they had survived. This is sufficient for the primary analysis of QALYs (i.e. area under EQ-5D curve) using Cox regression. The second row uses all 213 participants, as required for both the main 48-month cost-effectiveness analysis and our sensitivity analysis of QALYs using ANCOVA. ANCOVA relies more heavily than Cox regression on imputation: more than 25% of the EQ-5D scores at 36 months arose from multiplying imputed 'live' scores by estimated survival probabilities (see *Chapter 2*, *Design*).

As expected, the resulting mean quality of life declines faster over time for all participants (including those who have died) than for survivors only (see *Appendices 6.6* and *6.8*). Although at all time points except 1 month the mean is higher in the EUS group, none of the differences between EUS and non-EUS

TABLE 19 EQ-5D utilities by allocated group: unadjusted means (including deaths and imputed survivors)

Time of EQ-5D	Non-EUS group		EUS group		Total *n*	EUS minus non-EUS	
	n	Mean (SD)	*n*	Mean (SD)		Difference (95% CI)[a]	
Survivors (imputed to end of trial if necessary) and deaths before end of trial							
Baseline	106	0.801 (0.164)	107	0.807 (0.198)	213	0.007	(–0.042 to 0.056)
1 month	106	0.733 (0.285)	107	0.729 (0.265)	213	–0.005	(–0.079 to 0.070)
3 months	106	0.615 (0.305)	107	0.658 (0.289)	213	0.043	(–0.037 to 0.123)
6 months	106	0.535 (0.323)	107	0.550 (0.347)	213	0.015	(–0.076 to 0.106)
12 months	**106**	**0.449 (0.391)**	**107**	**0.509 (0.376)**	**213**	**0.061**	**(–0.043 to 0.164)**
18 months	87	0.377 (0.400)	90	0.394 (0.399)	177	0.017	(–0.102 to 0.135)
24 months	73	0.251 (0.347)	77	0.330 (0.392)	150	0.079	(–0.041 to 0.198)
36 months	44	0.189 (0.312)	47	0.226 (0.373)	91	0.037	(–0.106 to 0.181)
All participants (fully imputed beyond end of trial if necessary)[b]							
18 months	106	0.353 (0.387)	107	0.400 (0.382)	213	0.047	(–0.057 to 0.151)
24 months	106	0.260 (0.341)	107	0.328 (0.364)	213	0.068	(–0.027 to 0.163)
36 months	106	0.152 (0.260)	107	0.211 (0.306)	213	0.060	(–0.017 to 0.136)

SD, standard deviation.

a Confidence intervals for comparison only; to avoid multiple testing, analysis focuses on 12-month follow-up.

b Confidence intervals in this subtable, especially for 36 months, are narrower than in corresponding rows of previous subtable owing to increase in sample size.

groups is significant. *Table 19* reports imputed scores at 12 months as secondary outcomes, to reduce the chance of false-positive results due to multiple testing. By analogy, from the FACT data we chose only four more secondary outcomes – the corresponding AUC and 12-month scores of FACT-G and FACT-AC (see *Recruitment, participant flow and CONSORT*). For description rather than inference *Appendix 6.7* reports the corresponding raw data and basic *t*-tests at all time points, thus elaborating *Table 19* by including all eight quality-of-life measures shown in *Table 18* – at all seven follow-up points.

From death we scored EQ-VAS, FACT-G and its four subscales as zero, and FACT-AC (for which the worst score is 4.0) as 3.66, the worst reported within the COGNATE trial (see *Chapter 2*, *Statistical methods, Imputation and missing values*). The FACT scales and EQ-VAS then follow patterns similar to that of EQ-5D utilities in *Table 19*. By 36 months mean quality of life measured by FACT or EQ-5D had dropped below one-quarter of its optimum value. Although none of the 56 differences in *Appendix 6.7* is significant, all 56 are in favour of EUS. *Appendix 6.8* focuses on findings at 12 months from *Appendix 6.7*. The more powerful ANCOVA also shows no significant differences between intervention and control groups. *Appendices 6.9–6.11* show the models for EQ-5D, FACT-G and FACT-AC respectively – those suggested by the structural equation modelling (see *Refinement of quality-of-life measures*). *Table 20* shows the corresponding adjusted means and confidence intervals.

In the search for covariates, baseline measurements are highly significant for the two FACT summary measures and close to significance for EQ-5D. In all models the only other significant or near-significant covariates are biopsy T stage (one of the two covariates for survival) and age. Initial management plan is less significant in predicting quality of life than in predicting survival, especially for the two FACT models.

Table 21 shows the result of including interactions in the ANCOVA, illustrated by the same three participants used in *Table 17*, with low, medium and high baseline EQ-5D respectively.

Appendices 6.12–6.14 show the corresponding models for EQ-5D, FACT-G and FACT-AC respectively. At 12 months there was no significant difference in EQ-5D between EUS group and non-EUS group (see *Table 20* and *Appendix 6.9*). If we add interaction between EUS group and baseline EQ-5D to the model, both baseline EQ-5D and the interaction are significant (see *Table 21* and *Appendix 6.12*). Unlike survival, however, adding the interaction does not make the main effect of EUS group on EQ-5D significant. Replacing age by multimodal plan would again have little effect on the model, with management plan now a marginally significant predictor ($p = 0.048$), but only one of these two covariates is needed. *Table 21* shows that as baseline EQ-5D increases, the advantage of EUS decreases and is eventually reversed. Although an EUS participant's expected quality of life at 12 months is similar for the three examples, non-EUS participants' quality-of-life outcomes depend strongly on their initial EQ-5D.

Thus the analysis of both survival and EQ-5D shows highly significant interactions between participants' allocated treatment (whether they had drawn EUS in the COGNATE lottery) and baseline EQ-5D. EUS may benefit participants in poor health more than those whose health is not yet showing the worse effects of gastro-oesophageal cancer.

In planning the analogous analyses of covariance for the FACT-G (see *Appendix 6.13*) and FACT-AC (see *Appendix 6.14*), we predicted that the corresponding baseline FACT scores would also show significant interaction with allocated treatment. However those analyses showed that the baseline FACT

TABLE 20 Twelve-month quality of life by allocated group: means adjusted by ANCOVA without interaction

Scale	Non-EUS group	EUS group	EUS minus non-EUS (95% CI)	Significance
FACT-G[a]	2.15	2.27	0.12 (–0.27 to 0.51)	0.546
FACT-AC[b]	1.76	1.84	0.08 (–0.25 to 0.42)	0.624
EQ-5D utility[a]	0.444	0.503	0.060 (–0.041 to 0.161)	0.245

a Scores of zero are bad; positive differences favour EUS group.

b Scores of zero are good; positive differences favour non-EUS group.

TABLE 21 Twelve-month quality of life by allocated group: means adjusted by ANCOVA with interaction

Scale	Participant's baseline EQ-5D	Predicted score		Estimated difference (EUS minus non-EUS)
		Non-EUS group	EUS group	
EQ-5D at 12 months[a]	0.6	0.325	0.560	0.235
	0.8	0.483	0.543	0.060
	1.0	0.641	0.526	–0.115
FACT-G at 12 months[a]	0.6	1.509	2.372	0.863
	0.8	2.152	2.273	0.121
	1.0	2.875	2.182	–0.693
FACT-AC at 12 months[a,b]	0.6	1.986	1.721	–0.265
	0.8	1.755	1.841	0.086
	1.0	1.250	1.961	0.711

a Allowing for centre, pre-randomisation T stage and age.

b In FACT-AC, high scores are bad.

scores achieved no interactions. Instead we found that baseline EQ-5D surprisingly continued to interact strongly with allocated treatment. We interpret this as evidence that baseline EQ-5D may be better than other baseline measures at differentiating 'worse' patients (helped by EUS) from 'better' patients (not helped by EUS) in serious conditions.

Including interactions as well as covariates in the models predicting 12-month EQ-5D, FACT-G and FACT-AC had very little influence on the main effect of allocated treatment: the estimated differences between EUS and non-EUS groups when baseline EQ-5D takes its average value of 0.8 in *Table 21*, and the significance of those differences, are similar to the differences in *Table 20*.

Survival adjusted by EQ-5D (primary outcome)

Calculation, imputation and alternative analyses
Although unadjusted survival analyses used data from 221 participants, quality-adjusted survival used only the 213 participants with enough data on survival and quality of life (see *Recruitment, participant flow and CONSORT*; *Attrition, withdrawals and missing data*). An explanation of why we nevertheless based our primary analysis on this outcome is provided in *Chapter 2* (see *Design*; *Primary outcome measure: quality-adjusted survival*), measured in QALYs. All three measures of quality of life – EQ-5D, FACT-G and FACT-AC – include actual or imputed scale scores from all follow-up times when participants were known to be alive. Thus our primary analysis modified Cox regression to take account of variable length of follow-up while minimising the proportion of imputed quality-of-life scores.

However this method does not easily allow us to combine QALYs with costs in a cost-effectiveness analysis. For EQ-5D only, therefore, we carried out two additional analyses of quality-adjusted survival in which the AUC covers the same period for every participant – by truncating the AUC of the Cox regression at 12 months, and by extending it beyond the end of the trial to 48 months.

We followed all participants for at least 12 months unless they withdrew from the trial, although some missed isolated interviews. The first two steps in the stepped imputation process described in *Chapter 2* (see *Statistical Methods, Imputation and missing values*) results in a data set with complete survival, EQ-5D and FACT information up to 12 months, in which we had to impute survival for only three participants. Beyond 12 months the number of people with missing interviews or survival information increases rapidly, because no information at all is available beyond the end of the trial. The main Cox regression analysis uses imputed quality-of-life scores at later time points only for the subset of participants still in the trial at those time points.

For extra sensitivity analyses to 12 and 48 months, we estimated QALYs as the full area under the EQ-5D curve for every participant, and analysed these QALYs by standard ANCOVA like the preceding point estimates of quality of life. However the 12-month comparison of QALYs ignores both deaths and quality of life between 12 months and the end of the trial, while the 48-month comparison relies heavily on the few interviews that took place at 24 or 36 months.

In contrast, our primary outcome measure uses the area under the EQ-5D curve over observed periods of up to 48 months, and our survival analysis takes full account of censoring due to withdrawals and the end of the trial. Although quality-adjusted survival analysis is similar to standard survival analysis, it combines survival with EQ-5D and is therefore more dependent on baseline EQ-5D. Furthermore the interaction between baseline EQ-5D and allocated treatment is more significant.

Univariate comparisons (quality-adjusted survival curves and log-rank tests)
Participants in early tumour stages [tumour in situ (Tis), T1 or T2] before randomisation achieved significantly higher QALYs ($p = 0.008$ by univariate Mantel–Cox log-rank test), as did participants whose original management plan was not multimodal ($p = 0.002$). However allocated treatment, tumour type, age over 65 years, gender, tumour site and centre do not contribute to survival in isolation ($p = 0.09$,

0.20, 0.28, 0.44, 0.57 and 0.65 respectively). *Table 22* compares mean and median quality-adjusted survival between the two allocated treatments. The more robust median quality-adjusted survival times are 11 months for the control group and 13.5 months for the intervention group.

Figure 18 compares adjusted survival curves by allocated treatment. In contrast to the unadjusted survival curves, these two curves quickly separate, with consistently better experience in the intervention group than in the control group. Only 1 of the 10 people who survived beyond 4 years had maximum quality of life throughout their time in the trial. Similarly, fewer participants reached one, two or three QALYs in the trial than the corresponding numbers for unadjusted survival times.

Quality-adjusted survival in the three centres is shown in *Figure 19*. Although the 'other centres' curve is slightly above the curves for Gloucester and Aberdeen, it represents only 41 participants, and the difference could easily be explained by chance.

Survival adjusted by EQ-5D with multiple predictors
We undertook survival analyses by Cox regression for 213 participants for whom there was enough survival, FACT and EQ-5D data to estimate quality-adjusted survival; all contributed to the model estimates. As for unadjusted survival, our analysis censored observations at the time of withdrawal or loss to

TABLE 22 Mean and median QALYs by allocated group

Group	Mean QALYs (years, truncated at 4-year survival)		Median QALYs (years)	
	Mean	95% CI	Median	95% CI
Non-EUS	1.44	1.14 to 1.74	0.94	0.51 to 1.37
EUS	1.83	1.49 to 2.17	1.12	0.67 to 1.57
Both	1.65	1.42 to 1.88	1.05	0.76 to 1.34
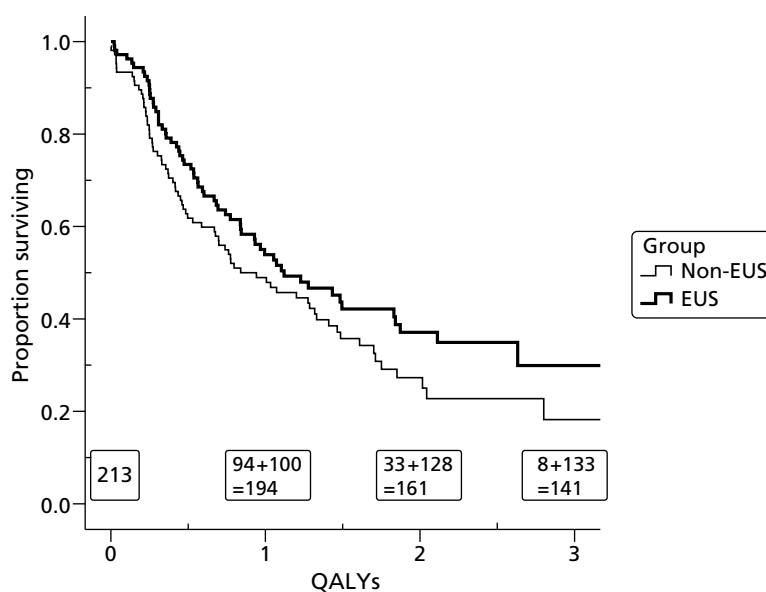
FIGURE 18 Quality-adjusted survival curves: EUS and non-EUS groups. Rectangular boxes above the *x*-axis record the number of participants available to estimate survival curve, subdivided into participants still alive plus cumulative deaths; thus these boxes exclude participants censored when the trial ended or they withdrew completely.
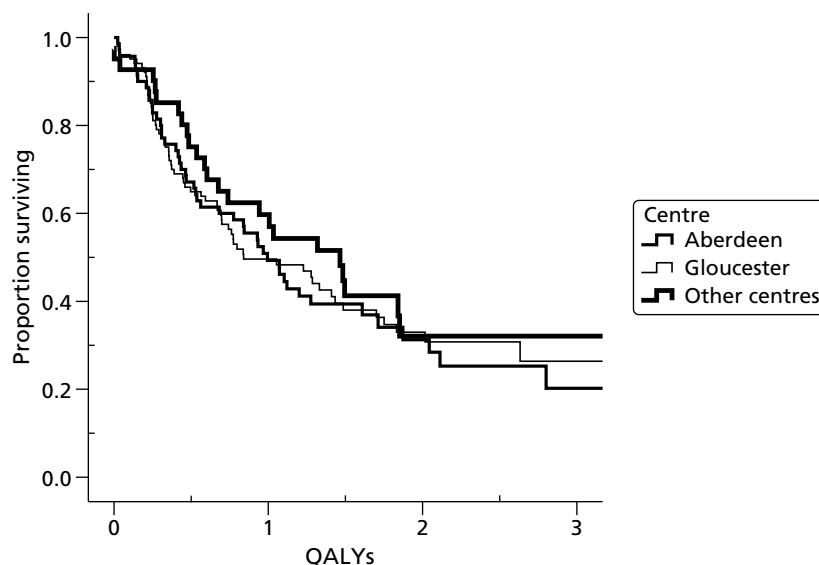
FIGURE 19 Quality-adjusted survival curves by centre.

follow-up. We fitted three models: the simplest fitted only the allocated treatment; the second model added the effects of centre and other significant or helpful covariates, but not interactions; and the third added the only significant interaction term.

In the model without covariates, the intervention group had better quality-adjusted survival than the control group, but not significantly better; the hazard ratio was 0.76 (95% CI from 0.54 to 1.06; $p = 0.106$). *Table 23* shows the second Cox regression, including covariates but not interaction. As before potential covariates were baseline characteristics: centre; baseline EQ-5D; aged over 65 years; gender; tumour site, pre-randomisation stage and type; and management plan before randomisation. In accordance with the analysis plan [see *Chapter 2*, *Statistical methods*, *Quality-adjusted survival (primary outcome) and survival*], the first two of these were automatically included. The other two selected – pre-randomisation T stage and multimodal plan – were the same as for unadjusted survival. However age, which replaced multimodal plan as a covariate for quality of life at 12 months, was not a significant predictor. The addition of these four covariates improved the contrast between intervention and control: the hazard ratio fell to 0.705 (95% CI from 0.499 to 0.995; $p = 0.047$), now significantly different from 1.0.

TABLE 23 Quality-adjusted survival by allocated group: Cox regression without interaction

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS	**0.705**	**0.499 to 0.995**	**0.047**
Other centres vs (Aberdeen or Gloucester)	1.043	0.711 to 1.531	0.828
Gloucester vs Aberdeen	0.871	0.693 to 1.095	0.236
(T3, T4, NR) vs (Tis, T1, T2)	**1.316**	**1.065 to 1.627**	**0.011**
Multimodal plan vs not	**1.399**	**1.133 to 1.728**	**0.002**
Baseline EQ-5D	0.445	0.180 to 1.098	0.079

Rows in bold type contribute significantly to the model.

TABLE 24 Quality-adjusted survival by allocated group: Cox regression including interaction

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS	0.712	0.504 to 1.005	0.054
Other centres vs (Aberdeen or Gloucester)	0.904	0.606 to 1.347	0.619
Gloucester vs Aberdeen	0.884	0.703 to 1.111	0.290
(T3, T4, NR) v (Tis, T1, T2)	1.232	0.993 to 1.529	0.058
Multimodal plan vs not	**1.417**	**1.146 to 1.751**	**0.001**
Baseline EQ-5D	**0.275**	**0.101 to 0.748**	**0.011**
Interaction: EUS and EQ-5D	**16.34**	**2.05 to 129.90**	**0.008**

Rows in bold type contribute significantly to the model.

Table 24 shows that interaction between baseline EQ-5D and allocated treatment in predicting QALYs is still highly significant ($p = 0.008$). The main effect of baseline EQ-5D also becomes significant ($p = 0.011$). Although the effect of EUS in the EUS group is no longer strictly significant, the hazard ratio 0.712 (95% CI from 0.504 to 1.005; $p = 0.054$) differs little from that without interaction. Although the initial management plan (and its covariates) had a large effect on quality-adjusted survival, it, like all other covariates except baseline EQ-5D, did not interact with the intervention. When the interaction was added, pre-randomisation T stage declined in importance as a predictor. As usual we kept centres in *Table 24* to show that that had no effect despite very different recruitment strategies.

Survival adjusted by FACT (secondary outcome)

Survival adjusted by FACT-G scale

We undertook another quality-adjusted survival analysis, again using 'AUC' but replacing EQ-5D with the more comprehensive FACT-G. For this measure a year of life with the best possible FACT score of 4.0 throughout gives a quality-adjusted survival of 1 year, and a year in which the average FACT score is 2.0 gives a quality-adjusted survival of only 6 months. Results were similar, but not identical, to those using
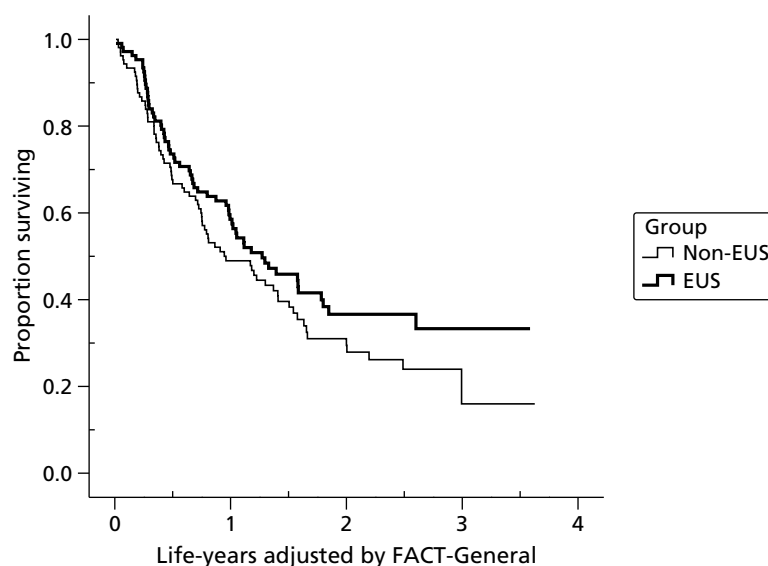
FIGURE 20 Survival curves adjusted by FACT-G by allocated treatment group.

TABLE 25 Survival adjusted by FACT-G: mean and median by allocated group

| Group | Mean QALYs (years) | | Median QALYs (years) | |
	Mean	95% CI	Median	95% CI
Non-EUS	1.46	1.19 to 1.72	0.94	0.55 to 1.34
EUS	1.78	1.49 to 2.06	1.27	0.79 to 1.76
Both	1.63	1.43 to 1.82	1.17	0.90 to 1.43

the conventional definition of quality-adjusted survival based on the EQ-5D. *Figure 20* shows the observed survival curves for the two groups. However a Mantel–Cox log-rank test shows no significant difference between the allocated treatment groups ($p = 0.11$).

Table 25 gives the corresponding means, medians and confidence intervals. The median survival times adjusted by FACT-G are 11.5 months for the control participants and 15 months for the intervention group. However Cox regression using allocated treatment group as single predictor did not show a significant effect of EUS: the hazard ratio was 0.774 with a 95% confidence interval from 0.550 to 1.089 ($p = 0.142$). Although adding management plan, pre-randomisation T stage, baseline FACT-G and centre to the covariates decreases the estimated hazard ratio to 0.72, and improves significance, the difference is still not significant (see *Appendix 6.15*).

We then analysed models with interactions. As for 12-month FACT-G scores, we considered interactions with baseline FACT-G and the usual covariates (see *Appendix 6.14*). As none of these were significant, we added baseline EQ-5D and its interaction with allocated treatment to the model (see *Appendix 6.15*). There was significant interaction with baseline EQ-5D but, as with the 12-month FACT-G, not with baseline FACT-G. Although baseline FACT-G still had a significant effect on the outcome, that effect was the same in both allocated treatment groups and does not affect the difference between them or the hazard ratio. As before, however, this hazard ratio applies only to participants with baseline EQ-5D equal to 0.8, the mean of the whole sample. Participants with lower baseline EQ-5Ds have an even smaller hazard ratio, whereas those with maximum baseline EQ-5Ds of 1.0 have hazard ratios close to 1.0. Thus EUS improves FACT-G-adjusted quality of life for the sickest participants but not the healthiest.

Quality-adjusted survival (adjusting for FACT-AC scale)

We undertook a third quality-adjusted survival analysis, again using 'AUC' but replacing EQ-5D with the FACT-AC scale, more focused on gastro-oesophageal cancer. *Table 26* shows that the median survival adjusted by FACT-AC was 12 months in the control group and 14 months in the intervention group, still not significantly different (log-rank test: $p = 0.16$). As usual there was no significant difference between centres ($p = 0.53$). Cox regression without covariates has an estimated hazard ratio of 0.78. Adding pre-randomisation T stage (T2 or better vs T3 or worse), management plan (multimodal vs not) and baseline FACT-AC as covariates again contributed significantly to the model, and reduced the hazard ratio for EUS to 0.77 (95% CI from 0.54 to 1.08; $p = 0.137$). Adding the interaction between allocated treatment and baseline EQ-5D to the other covariates is again highly significant ($p = 0.006$) but does not enhance

TABLE 26 Survival adjusted by FACT-AC: mean and median by allocated group

| Group | Mean QALYs (years) | | Median QALYs (years) | |
	Mean	95% CI	Median	95% CI
Non-EUS	1.51	1.22 to 1.81	1.01	0.61 to 1.40
EUS	1.71	1.43 to 1.99	1.15	0.69 to 1.62
Overall	1.68	1.46 to 1.90	1.13	0.90 to 1.36

the main effect of EUS (hazard ratio 0.79; 95% CI from 0.56 to 1.11). Although this interaction reduces the contribution of pre-randomisation T stage to the model, that of multimodal plans remains highly significant ($p = 0.001$).

Quality-adjusted survival: sensitivity analyses with fully imputed data

Table 27 displays means and confidence intervals from comparisons of QALYs fully imputed to 1 or 4 years for all 213 participants, both before and after ANCOVA. The similarity between the estimated differences in the top and bottom of the table confirms that the conclusions of our comparative analyses are not sensitive to choice of method or covariates.

Before ANCOVA, the cost-effectiveness analysis (see *Health economics*) used these fully imputed estimates in its own analysis to test whether those conclusions are sensitive to the choice of whether or not to discount benefits as well as costs. In this way, sensitivity analyses in this section analysing effectiveness and in the subsection analysing cost-effectiveness (see *Health economics; Sensitivity analysis at 48 months with undiscounted quality-adjusted life-years*) have used precisely the same data.

Quality-adjusted survival to 12 months

Over the first 12 months of the trial, the mean quality-adjusted survival is only 0.03 years (11 days) longer in the intervention group. This estimate is consistent with the mean differences in EQ-5D of between −0.005 and 0.061 (see *Table 19*) over those 12 months, and the similar proportions of the two groups surviving to 12 months (70% and 67% in intervention and control groups respectively). As the only significant covariate is baseline EQ-5D ($p = 0.001$), the lower half of *Table 27* adjusts for only that and centre ($p = 0.633$). This hardly changes the effect of EUS on quality-adjusted survival. The only interaction that adds significantly to this model is with baseline EQ-5D itself, which also increases the significance of the main EQ-5D effect ($p < 0.001$). As before, the effect of EUS is larger for low initial quality of life. However this does not change estimates or non-significance.

Quality-adjusted survival to 48 months

The mean quality-adjusted survival over 48 months is 0.209 years (76 days) longer in intervention than control group; the medians are 1.102 and 0.784 years respectively – a difference of 0.318 years (116 days). The difference in medians is larger than in the corresponding survival analysis (see *Table 22*), but the difference in means is smaller. The fourth row of *Table 27* adds covariates to the comparison, and *Appendix 6.17* gives the full model. The same significant covariates – multimodal plan, pre-randomisation T stage and baseline EQ-5D – appear in this model as in the primary Cox regression analysis, again improving significance. Adding the usual interaction with baseline EQ-5D improves it further (see *Appendix 6.18*). Unlike the Cox regression, however, this model does not show significant difference between allocated treatments.

TABLE 27 Sensitivity analyses of fully imputed quality-adjusted survival by allocated group: raw means and estimated marginal means from ANCOVA without interactions

Time period	Non-EUS ($n = 106$)	EUS ($n = 107$)	EUS minus non-EUS (95% CI)	Significance
Mean QALYs (EQ-5D AUC)				
Not allowing for covariates				
To 12 months	0.5481	0.5785	0.0304 (−0.0422 to 0.1029)	0.410
To 48 months	1.2012	1.4098	0.2086 (−0.0720 to 0.4892)	0.144
Allowing for covariates				
To 12 months	0.5555	0.5839	0.0284 (−0.0424 to 0.0993)	0.430
To 48 months	1.0903	1.3007	0.2103 (−0.0578 to 0.4784)	0.123

TABLE 28 Treatment plan at randomisation by plan amended after EUS

Group	Treatment plan at randomisation	Treatment plan as amended after EUS or post-randomisation MDT				
		EMR	Surgery	Neo-adjuvant chemotherapy before surgery	Multimodal including palliative	Total
EUS	EMR	0	1	0	0	**1**
	Surgery	2	10	1	5	**18**
	Neo-adjuvant chemotherapy before surgery	4	7	49	7	**67**
	Multimodal	0	2	0	21	**23**
	Total	**6**	**20**	**50**	**33**	**109**
Non-EUS	EMR	3	1	0	1	**5**
	Surgery	0	9	0	1	**10**
	Neo-adjuvant chemotherapy before surgery	2	6	55	13	**76**
	Multimodal	0	1	2	16	**19**
	Total	**5**	**17**	**57**	**31**	**110**

Shaded cells represent changes of plan.

Change in management plans (secondary outcome)

Having found that EUS improves quality-adjusted survival, especially for those with low quality of life at baseline, we explore the mechanism of this effect by considering changes in management plans. *Table 28* compares plans at the time of randomisation into the trial with those recorded after EUS (or its alternative in the control group), but before treatment started; the diagonal, unshaded cells represent no change.

Table 28 shows no difference between intervention and control groups in the number of changes from initial, pre-randomisation management plan to updated plan: about one-quarter of each group changed – 29 out of 109 in the intervention group compared with 27 out of 110 in the control group. Given the improved survival and quality-adjusted survival, we did not expect this similarity. Hence we scrutinised the two subgroups who changed plans, although conscious that the small numbers in both meant that even large differences could not be statistically significant. Over 60% of initial plans in each group were for neo-adjuvant chemotherapy before surgery, with very few planned EMRs. Most changes in each group were from neo-adjuvant therapy before surgery to another option. In the intervention group 61% (11/18) of these changes (in the third row of *Table 28*) were to therapeutic treatment (EMR or surgery), compared with only 38% (8/21) in the control group. Among all changes of plan, more in the intervention group changed to a more optimistic plan (those below rather than above the leading diagonal of *Table 28*) than in the control group [52% (15/29) vs 41% (11/27)]. Half of the changes in the intervention group were recorded, explicitly or implicitly, as due to EUS results. Some changes in the control group cited other staging tests done after randomisation, typically PET, laparoscopy or CT.

Furthermore changes in plans improved survival more in the intervention group than in the control group. As we expected, those whose final plan was multimodal had significantly worse survival in both groups (see *Figure 16*). However this difference was more marked in the intervention group than in the control group (mean survival shorter by 354 and 205 days respectively). *Table 29* shows that this contrast between groups stemmed from those whose plans changed: in the intervention group mean survival was 650 days (namely 452 vs 1102) shorter in those who changed to multimodal than in other changers, but in the control group the difference was only 58 days (namely 564 vs 622). For those with unchanged plans the

TABLE 29 Survival in days by whether participants' management plans changed

Group	Final plan	Not multimodal Mean (median)	n	Multimodal Mean (median)	n	Total Mean (median)	n
EUS	Changed	1102[a]	17	452 (268)	12	847 (800)	29
	Unchanged	858 (811)	59	605 (478)	21	794 (657)	80
All EUS		914 (940)	76	560 (452)	33	808 (717)	109
Non-EUS	Changed	622 (664)	12	564 (419)	15	623 (633)	27
	Unchanged	760 (661)	67	455 (406)	16	708 (596)	82
All non-EUS		760 (664)	79	555 (406)	31	704 (664)	110
Both groups	Changed	986[a]	29	517 (419)	27	787 (664)	56
	Unchanged	804 (783)	126	546 (440)	37	749 (640)	163
Total		833 (800)	155	558 (421)	64	755 (640)	219[b]

a As most survived to the end of data collection, we cannot estimate the median.
b Two participants could not contribute to survival analysis because they withdrew before the first death. Two of the 221 who did contribute to survival analysis had no recorded final plan.

corresponding survival differences were close – 253 days (namely 605 vs 858) in the intervention group and 305 days (namely 455 vs 760) in the control group.

Although none of these differences was significant, they are consistent with EUS identifying participants whose original management plans were inappropriate, especially those originally scheduled for multimodal treatment but able to benefit from surgery, and those originally scheduled for surgery despite poor prognosis (see *Chapter 1*, *Literature review*, *Treatment*). In contrast those who changed plan in the control group experienced intermediate survival, arguably because they were on the borderline between alternative plans.

Table 30, restricted to 145 participants from the primary analysis who survived to 12 months (with the result that four of its eight cells contain fewer than 10 participants), illustrates how EUS also improved quality of life. All four sets of survivors who had changed plan (namely intervention or not by whether final plan was multimodal or not) reported higher mean EQ-5D scores than the corresponding set of survivors whose plan remained unchanged, and all four sets of survivors in the intervention group (namely plan changed or not by whether final plan was multimodal or not) reported higher mean EQ-5D scores than the corresponding set of control group survivors. When we extended *Table 30* to all 213 participants in the primary analysis by ascribing zero quality of life to those who did not survive to 12 months, we found similar benefits from EUS: 17 intervention participants who changed to non-multimodal plans had mean EQ-5D of 0.642; 16 control participants with confirmed multimodal plans had mean EQ-5D of 0.284, and the other six subgroups lay between these two extremes.

Table 31, restricted to the 213 participants for whom we could estimate costs, illustrates how EUS reduced costs. Control participants who changed plan in either direction incurred intermediate mean costs – above confirmed multimodal plans but below confirmed non-multimodal plans. In contrast intervention group participants who changed plan in either direction incurred lower costs than intervention group participants with confirmed plans of either type. That these findings are analogous to those of *Table 29* is not surprising, because cost generally depends on length of survival. However the relationship is not

TABLE 30 EQ-5D in survivors at 12 months by whether participants' management plans changed

Group	Final plan	Not multimodal Mean (median)	n	Multimodal Mean (median)	n	Total Mean (median)	n
EUS	Changed	0.839 (0.838)	13	0.848 (0.816)	6	0.842 (0.819)	19
	Unchanged	0.703 (0.726)	42	0.691 (0.759)	13	0.700 (0.759)	55
All EUS		0.735 (0.760)	55	0.740 (0.796)	19	0.736 (0.760)	74
Non-EUS	Changed	0.731 (0.708)	8	0.722 (0.700)	8	0.726 (0.707)	16
	Unchanged	0.683 (0.760)	46	0.504 (0.649)	9	0.653 (0.760)	55
All non-EUS		0.690 (0.760)	54	0.607 (0.725)	17	0.670 (0.725)	71
Both groups	Changed	0.798 (0.760)	21	0.776 (0.778)	14	0.789 (0.760)	35
	Unchanged	0.692 (0.760)	88	0.615 (0.691)	22	0.677 (0.759)	110
Total		**0.712 (0.760)**	**109**	**0.677 (0.741)**	**36**	**0.704 (0.760)**	**145**

TABLE 31 Costs over 48 months by whether participants' management plans changed

Group	Final plan	Not multimodal Mean (median)	n	Multimodal Mean (median)	n	Total Mean (median)	n
EUS	Changed	23.7 (23.8)	17	24.7 (21.1)	12	24.1 (22.1)	29
	Unchanged	32.9 (30.2)	58	25.9 (23.7)	20	31.1 (29.4)	78
All EUS		30.8 (28.2)	75	25.4 (22.1)	32	29.2 (27.0)	107
Non-EUS	Changed	27.2 (20.6)	11	23.8 (22.8)	14	25.3 (22.7)	25
	Unchanged	37.5 (34.2)	65	20.5 (13.1)	16	24.1 (30.6)	81
All non-EUS		36.0 (33.0)	76	22.0 (20.0)	30	32.0 (27.9)	106
Both groups	Changed	25.1 (22.9)	28	24.2 (22.4)	26	24.6 (22.4)	54
	Unchanged	35.5 (33.0)	122	23.5 (21.3)	36	32.8 (29.8)	158
Total		**33.4 (29.6)**	**151**	**23.8 (22.1)**	**62**	**30.6 (27.4)**	**213**

All costs are in £, to the nearest 10p.

straightforward, as costs are also subject to other influences; for example, long-term survivors avoid the typically high cost of terminal care.

In summary, *Tables 29–31* are exploratory tables, designed to explain the apparent inconsistency between major differences in survival and quality of life between intervention and control groups yet similar numbers of changes in management plans. Unlike the definitive analyses of outcome, they do not seek to adjust for covariates, even though the groups are not comparable: instead they use raw means and medians to estimate utility and cost, and simple univariate Kaplan–Meier analyses to estimate survival. Although changes of management plan, and the resulting final plan, are themselves trial outcomes

TABLE 32 Treatment plan by actual treatment

| Group | Treatment plan at randomisation | Treatment delivered | | | | |
		EMR	Surgery	Neo-adjuvant chemotherapy before surgery	Multimodal including palliative	Total
EUS	EMR	0	0	0	1	**1**
	Surgery	2	10	1	5	**18**
	Neo-adjuvant chemotherapy before surgery	4	8	36	19	**67**
	Multimodal	1	3	1	18	**23**
	Total	**7**	**22**	**38**	**43**	**109**
Non-EUS	EMR	2	1	0	2	**5**
	Surgery	0	9	0	1	**10**
	Neo-adjuvant chemotherapy before surgery	2	6	43	25	**76**
	Multimodal	1	1	2	15	**19**
	Total	**5**	**17**	**45**	**43**	**110**

Shaded cells represent changes between plan and treatment.

dependent on covariates, traditional adjustment by ANCOVA would obscure the relationships we seek to understand.

Table 32 shows that treatments actually delivered did not always coincide with management plans at randomisation. In particular several participants in both groups started neo-adjuvant chemotherapy but did not progress to surgery, usually because the tumour remained unresectable. Their actual treatment was therefore multimodal. Again there was no significant difference between the pattern of change in the intervention and control groups. However *Figure 21* shows considerable differences between centres, both in amendments to the initial treatment plan and in the actual treatment delivered. In both Aberdeen and Gloucester some 30% of participants were originally scheduled for neo-adjuvant chemotherapy followed by surgery but were switched from it. However most changes in Gloucester occurred before the updated plan was confirmed, while in Aberdeen such changes occurred mostly between the updated plan and actual treatment. Other centres changed less often than Aberdeen and Gloucester, and ended with fewer participants receiving only multimodal treatment, as one might expect from their better initial tumour T stage (see *Table 10*) and baseline quality of life (see *Table 25*).

Other secondary outcome measures

Complete resection
Table 33 shows that the proportion of complete resections among curative operations (including those after neo-adjuvant chemotherapy) was 91% in the intervention group but 79% in the control group ($p = 0.085$). The relative risk of incomplete resection or no attempt (often described as 'open and close' surgery, which happened only once) in the intervention group was 0.44 with a 95% confidence interval from 0.17 to 1.17. Thus, although the two groups did not differ significantly, control group operations were more than twice as likely to be incomplete.

Table 34 shows that, although the complete resection rate was highest in Gloucester, there was no significant difference between centres ($p = 0.30$). Aberdeen and Gloucester had relatively more
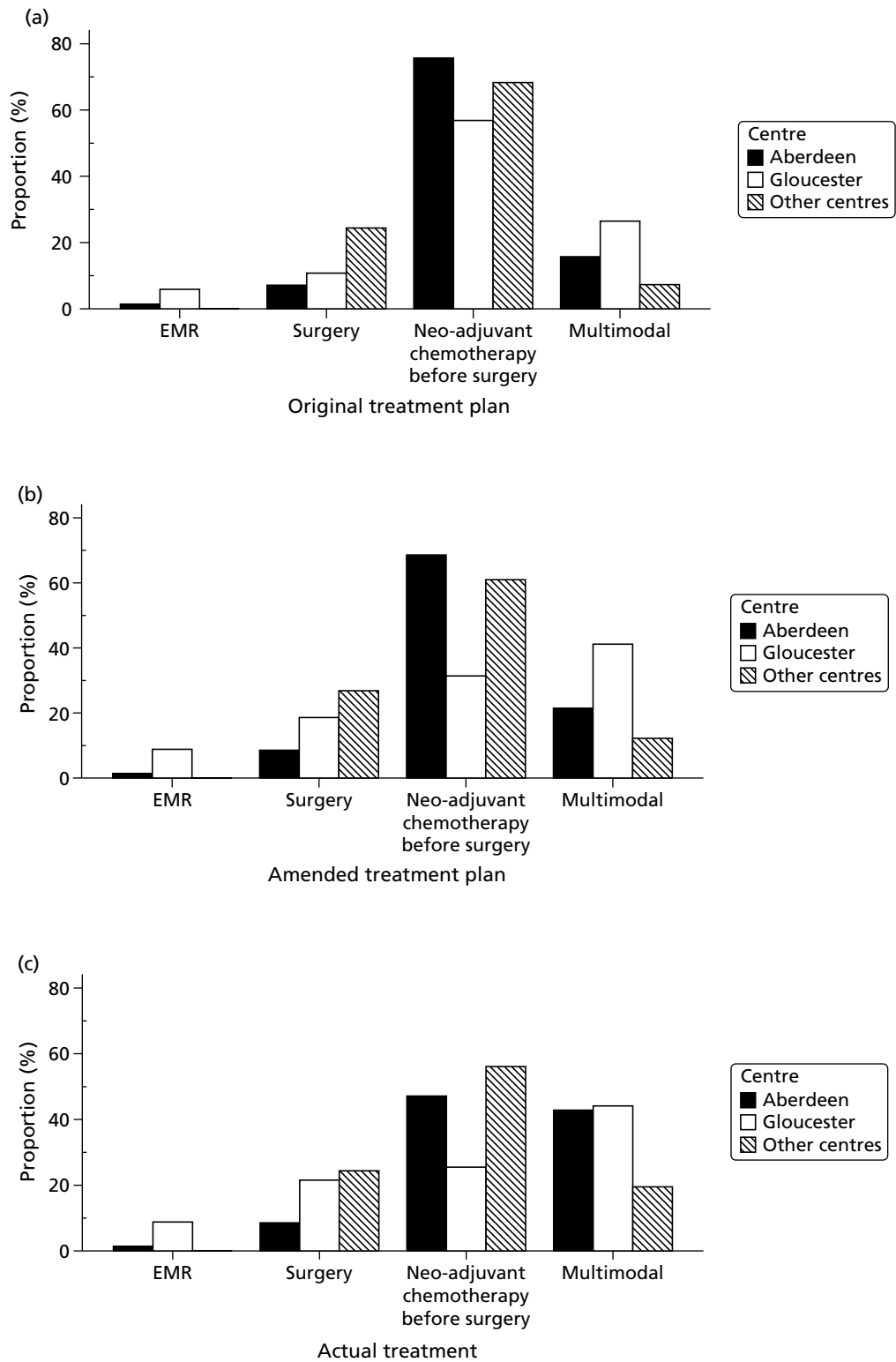
FIGURE 21 Management plan changes and treatment received by centre (three groups).

TABLE 33 Complete resection by allocated group among those having surgery

	Group					
Resection	Non-EUS: no. (%)		EUS: no. (%)		Total: no. (%)	
Complete	44	(79)	48	(91)	92	(84)
Incomplete	11	(20)	5	(9)	16	(15)
No attempt	1	(2)	0	(0)	1	(1)
Total	**56**	**(100)**	**53**	**(100)**	**109**	**(100)**

TABLE 34 Complete resection by centre among those having surgery

	Centre					
Resection	Aberdeen: no. (%)		Gloucester: no. (%)		Other Centres: no. (%)	
Complete	28	(78)	38	(90)	26	(84)
Incomplete	8	(22)	4	(10)	4	(13)
No attempt	0	(0)	0	(0)	1	(3)
Total	**36**	**(100)**	**42**	**(100)**	**31**	**(100)**

TABLE 35 Cause of death

	Group					
Cause of death	Non-EUS: no. (%)		EUS: no. (%)		Total: no. (%)	
Alive at end of trial	48	(44)	58	(54)	106	(49)
Cancer	47	(43)	31	(28)	78	(36)
Other but cancer contributed	5	(5)	15	(14)	20	(9)
Unrelated to cancer	2	(2)	1	(1)	3	(1)
Death related to treatment (not including EUS)	6	(5)	4	(4)	10	(5)
Total	**108**	**(100)**	**109**	**(100)**	**217**	**(100)**

non-surgery participants than other centres: the proportion of complete resections across the whole sample was 40% (28/70) in Aberdeen, 35% (38/109) in Gloucester, and 62% (26/42) elsewhere. Of those who changed to a non-multimodal plan and subsequently proceeded to surgery, two in the control group, but none in the intervention group, had incomplete resections.

Cause of death and metastases

We asked all participating centres to report all complications, including deaths within 30 days, which could possibly have arisen from EUS scans – with the intention of invoking our SOP for safety monitoring[158] for any reported. As centres reported no such complications or adverse events associated with EUS we did not need to invoke that SOP. However centres considered 10 deaths to be associated with complications of surgery or chemoradiotherapy. Only three deaths were reportedly unrelated to cancer (*Table 35*). The proportion of deaths with cancer as main cause was significantly higher in the control group (78%) than in the intervention group (61%). The main cause was cancer in 29 out of 45 deaths (64%) in Aberdeen, 32 out of 44 deaths (73%) in Gloucester, and 17 out of 22 deaths (77%) in the other centres; this difference was not significant. Two of the six treatment-related deaths in the control group occurred after a changed plan (one from multimodal to neo-adjuvant surgery, and one from neo-adjuvant to immediate surgery).

TABLE 36 Disease status at end of trial or death by allocated group

	Group					
Status	**Non-EUS: no. (%)**		**EUS: no. (%)**		**Total: no. (%)**	
Unknown or missing	37	(34)	43	(39)	80	(36)
No disease	12	(11)	10	(9)	22	(10)
Local tumour	30	(27)	25	(22)	55	(25)
Metastases only	8	(7)	4	(4)	12	(5)
Local tumour + metastases	23	(21)	29	(26)	52	(24)
Total	**110**	**(100)**	**111**	**(100)**	**221[a]**	**(100)**

a Excluding two very early withdrawals.

At the end of the trial or death, 63 people were known to have metastases, with approximately equal numbers in EUS and non-EUS groups. However this information was not available for over one-third of participants in each group. Although *Table 36* documents all available final disease stages, follow-up varied between 1 and 4 years to achieve the best estimates of survival, a principal objective. Hence our data on final disease stage mean little, and we could not have done better without recalling survivors to hospital at the end of the trial for intrusive tests and conducting post-mortems on all who died. Instead we inferred that those who survived without adverse comment in their clinical report form (CRF) or medical record were disease free.

Analysis by 'treatment received'
Fourteen participants allocated to EUS did not receive it, for varied reasons: early withdrawal (excluded from most analyses) or failure to attend accounted for six; appointments could not be arranged in time for another three; and equipment failed twice. Three allocated to the control group did receive EUS, only two at the request of medical staff. As a form of sensitivity analysis we repeated our main analyses – survival with and without adjusting for quality of life, changes in management plan and complete resection – by 'EUS received'.

Management plan changes and complete resection
None of the three participants who received an EUS to which they were not randomised changed plan. One received multimodal treatment as planned, and the other two were completely resected during surgery after neo-adjuvant chemotherapy.

Two of the fourteen participants who did not receive an EUS to which they were randomised withdrew too soon for management plan changes to be recorded. Three of the remaining twelve changed plan – one from surgery to multimodal, two from neo-adjuvant therapy to multimodal; a further two whose plan was surgery after neo-adjuvant therapy actually received multimodal treatment, and one was switched from multimodal to surgery. Thus analysis by 'treatment received' still has similar proportions of plan changes in EUS and non-EUS groups.

Only one of the 14 received surgery, and was completely resected. Thus 52 of those who received EUS, and 57 of those who did not, had surgery; 47 (90%) and 45 (79%) of these respectively were completely resected. While these proportions were almost identical to those in the main analysis, the intervention and control resection rates differed more in the 'treatment-received' analysis: 47% of those who received EUS, and 37% of those who did not, were completely resected. However this difference was still not significant.

TABLE 37 (Quality-adjusted) survival: analyses by 'treatment allocated' and 'treatment received'

Analysis	EUS effect (no interaction)		EUS effect (model with interaction)		Interaction between EUS and baseline EQ-5D	
	Hazard ratio	Significance	Hazard ratio	Significance	Hazard ratio	Significance
Survival						
Treatment allocated	0.706	0.048	0.704	0.046	13.18	0.016
Treatment received	0.697	0.041	0.693	0.041	7.40	0.053
QALY (EQ-5D)						
Treatment allocated	0.705	0.047	0.712	0.054	16.34	0.008
Treatment received	0.711	0.055	0.718	0.065	5.20	0.092
QALY (FACT-G)						
Treatment allocated	0.724	0.065	0.729	0.073	14.94	0.012
Treatment received	0.698	0.043	0.704	0.050	6.43	0.072
QALY (FACT-AC)						
Treatment allocated	0.769	0.134	0.785	0.170	19.31	0.008
Treatment received	0.760	0.122	0.766	0.138	7.41	0.060

Cox regression survival analyses

The general pattern of Cox regression models did not change (see *Appendices 6.19–6.22*, summarised in *Table 37*). In particular the estimated main effect of EUS after allowing for baseline quality of life, centre, management plan and initial T stage was very similar in size and significance, with marginally stronger effects than in the main analyses by treatment allocated. However the interaction with baseline EQ-5D was less marked, and significance level was between 5% and 10% for all four outcome measures: survival, and survival adjusted by EQ-5D, FACT-G and FACT-AC respectively. Nevertheless the size and significance of the main effects was unchanged.

Other analyses

Our models have consistently shown that, apart from random allocation to EUS, the main prognostic indicators for both survival and quality of life were pre-randomisation tumour stage, planned multimodal treatment, baseline quality of life, and the interaction between that baseline and random allocation to EUS. In particular there is no hint in any of our many multivariate analyses of different types that centre, tumour site or tumour type had any influence on outcome.

Palser *et al.*[5] have strongly recommended EUS for oesophageal tumours but not for gastric tumours. As none of our extensive analyses provides any support for this, in particular any evidence of interaction between allocated treatment and tumour site, we offer two case studies of patients with gastric cancer to illustrate our analyses:

> *A patient aged 61 with adenocarcinoma of the stomach was randomised after screening barium study, chest radiography, CT scan and endoscopy. The management plan of neo-adjuvant chemotherapy plus surgery was changed to multimodal treatment after endoscopic ultrasound revealed 'a bulky, locally advanced tumour unsuitable for resection'. Staging at randomisation was T3/T4 N1 but endoscopic ultrasound confirmed that staging was T4 N1. The patient died 7 months after randomisation of cancer as the main cause with local tumour only.*

A patient aged 70 with gastric adenocarcinoma was randomised in October 2007 to the EUS arm of COGNATE after endoscopy and chest radiography. The earlier biopsy report was 'highly suspicious for but not diagnostic of invasive adenocarcinoma of intestinal type'. The original management plan of surgery was changed to multimodal treatment as 'Endoscopic ultrasound showed a T3 lesion; patient therefore not for surgery'. Pre-randomisation staging was T2/T3 and endoscopic ultrasound confirmed T3 N1. The patient, who was frail with a lot of comorbidities, opted for palliative care and died 8 months later; cause of death was unrelated to cancer, with local tumour only.

Who might benefit from endoscopic ultrasound?

We had hoped to estimate the general proportion of gastro-oesophageal cancer patients who benefit from EUS. However differences between our centres (see *Figure 4*) and our unsuccessful attempts to recruit more than two centres unequivocally in equipoise prevented us from estimating that proportion with any confidence.

Therefore restricting this estimation to the COGNATE study population in *Figure 4*, we see that, of the 1124 patients with positive histology, 567 were potentially eligible for the trial and therefore for EUS. However the 114 not consenting to the trial could still potentially benefit from EUS. Of the 453 who did consent, eight were no longer eligible by the time of the MDM, and eight more proved unsuitable for EUS. At first sight, therefore, we would expect $(567 \times 437/453) = 547$ (49% of 1124) patients with gastro-oesophageal cancer to benefit from EUS in the way shown by the trial. To illustrate the differences between centres, we estimate by subdividing *Figure 4* that this proportion varies from 70/172 (41%) in Aberdeen to 101/163 (62%) in Gloucester.

However the interaction between EUS group and baseline EQ-5D shows those with the best initial quality of life do not benefit from EUS. Estimating from our interaction models that about one-quarter may not benefit, we predict that only 410 (36%) of the original 1124 were likely to benefit. In short this trial suggests that, by undertaking EUS for about 48% of patients with histologically proven gastro-oesophageal cancer, one can benefit about 36%.

Health economics

Unit costs

As we finished data collection in July 2009, we chose to use unit costs from 2008 for all our analyses. These covered two time periods: up to 12 months after randomisation and up to 48 months after, by which time many costs and QALYs had inevitably been censored. *Table 38* shows these unit costs and their sources.

Missing cost and quality-adjusted life-year data

Table 39 shows numbers of participants with partially or completely missing data over time periods. We pursued missing data throughout the trial (see *Chapter 2*, *Design*, *Electronic data collection and storage*). Our electronic database alerted researchers interviewing participants to missed questions and overdue interviews, and verified data quality. Researchers collecting and entering cost data from hospital records also received database prompts, and regular checks from the trial centre. By these means over the first 12 months we collected costs and survival data for all but three participants who withdrew early.

Frequency of resource use

Table 40 (full table in *Appendix 7.1*) summarises the mean frequencies of resource use for secondary care contacts and hospital prescribed drugs over six time periods for the 213 participants in the primary analysis. This information comprises data put into the COGNATE database via laptop computers by research professionals at each study site. The effective sample throughout comprises those still in the trial, namely survivors observed throughout the period, and those who died before the end of the period and before 31 July 2009. Participants, dead or alive, for whom the end of the trial occurred during that period, get half the weight of those in the trial for the whole of that period.

TABLE 38 Unit costs of health service use for 2008 with source[a]

Health-care resource	Unit	Unit cost (£)	Details and source
Hospital outpatient clinic	Consultation	Various	Costed by specialty[b]
Tests and investigations	Procedure	Various	Costed by procedure[b]
Day surgery	Procedure	Various	Costed by procedure[b]
Hospital inpatient episode	Bed-day	Various	Costed by procedure[b]
Chemotherapy	Cycle	Various	Costed by setting[b]
Radiotherapy	Course	Various	Costed by setting and fraction[b]
Prescribing	Item	Various	Costed by BNF entry[c]
Stent	Unit	£1222	2005 cost[d] updated to 2008 using the Hospital and Community Health Services Pay and Price Index[e]
PAMs, e.g. dietitian, physiotherapist	Consultation	Various	Costed by profession[e]
PET scan	Scan	£1000	From Aberdeen Royal Infirmary Nuclear Medicine and Finance Departments[f]

BNF, *British National Formulary*; PAMS, Professionals Allied to Medicine.

a NHS costs to nearest £, including salary, employers' costs, overheads and capital costs.

b From NHS Reference Costs 2007–8.[152]

c From Prescription Cost Analysis NHS.[148]

d From Shenfine 2005[154]

e From Curtis 2008.[155]

f Including personal communication from Dr Howard Gemmell.

TABLE 39 Number of missing and censored QALYs and costs by time period

No. of participants:	Time period (months)					
	0–12	12–18	18–24	24–36	36–48	48–60[a]
Already dead	0	68	94	111	129	*(133)*
With full costs and QALYs	118	84	59	37	0	*(0)*
With QALYs partly imputed	92[b]	38	23	13	14	*(0)*
Costs and QALYs partly imputed	3	20	14	15	18	*(10)*
Costs and QALYs fully imputed	0	3	23	37	52	*(70)*
Percentage with cost imputed (partially or fully)	1	11	17	24	33	N.A.[a]

N.A., not applicable.

a We did not use this interval in economic analysis or the associated sensitivity analysis of imputed QALYs in *Table 27* (see *Chapter 2, Statistical methods, Imputation and missing values*).

b To estimate number of QALYs gained between 0 and 12 months combines survival times with EQ-5D utilities from five questionnaires. All but 17 of these 92 had at least three of the five EQ-5D utilities, including zeros after death.

Hospital inpatient stay was considerably shorter in the EUS group (21.6 days) than in the non-EUS group (29.9 days). Almost all these stays were cancer related (20.8 and 29.5 days respectively). Both patterns were also present in the first year of the trial. In contrast outpatient visits were slightly more frequent in the EUS group than in the non-EUS group over 48 months but not over 12 months.

TABLE 40 Estimated mean frequency of contacts with secondary care health services and drugs use by 213 patients up to 60 months

Time period (months)[a]	EUS group						Estimated mean frequency for complete case	Non-EUS group		
	Estimated mean frequency in complete interval							Estimated mean frequency in complete interval		
	0–12	12–18	18–24	24–36	36–48	48–60		0–12	12–18	18–24
Effective sample size n[b]	106.5	97.5	83	63	33.5	10		105	96	80.5
Outpatient visits	11.31	2.47	2.90	2.97	1.28	0.80	**21.73**	11.29	2.06	1.45
Inpatient stay for any cause (no of bed-days)	13.28	2.91	1.80	3.05	0.57	0.00	**21.60**	18.00	2.02	3.23
EUS as day case (no. of day cases)	0.90	0.00	0.00	0.00	0.00	0.00	**0.90**	0.03	0.00	0.00
Surgery (count, no. of bed-days)	0.61, 13.68	0.00	0.00	0.00	0.00	0.00	**0.61, 13.68**	0.62, 15.74	0.00	0.00
EMR as day case	0.08	0.02	0.01	0.00	0.00	0.00	**0.11**	0.08	0.01	0.00
EMR as inpatient (count, no. of bed-days)	0.03, 0.06	0.00	0.00	0.00	0.00	0.00	**0.03, 0.06**	0.01, 0.03	0.00	0.00
Chemotherapy (no. of cycles)	2.24	0.25	0.35	0.33	0.00	0.00	**3.17**	2.82	0.14	0.26
Radiotherapy (no. of fractions)	5.44	0.35	0.48	0.17	0.00	0.00	**6.44**	5.35	0.54	0.17
Hospital prescribed drugs except for chemotherapy or surgery (no. of items)	28.93	5.26	4.67	5.40	3.01	5.90	**53.18**	31.79	4.31	4.09

This table is the first step to unbiased point estimates of mean costs over 4 years; as it has no single sample size or easy estimate of standard errors, however, our analysis of uncertainty will focus on bootstrapping mean costs after imputing those costs censored at the end of data collection.

a 0–12 months is the first year after randomisation; 12–18 months is the period between 12 and 18 months after randomisation, etc.

b The sample size decreases to include only those who were still in the trial: survivors observed throughout the interval, and those who died before the end of the interval but were randomised early enough for the end of the interval to occur before 31 July 2009. Participants (dead or alive) for whom the end of the trial occurred during the interval were included but given half of the weight of those in the trial for the full interval.

			Estimated mean frequency for complete case	Total sample							Estimated mean frequency for complete case
				Estimated mean frequency in complete interval							
24–36	36–48	48–60		0–12	12–18	18–24	24–36	36–48	48–60		
60	31	10		211.5	193.5	163.5	123	64.5	20		
1.75	1.16	0.00	**17.71**	11.30	2.27	2.19	2.37	1.22	0.40		**19.75**
3.47	3.16	0.00	**29.88**	15.62	2.47	2.50	3.25	1.81	0.00		**25.66**
0.00	0.00	0.00	**0.03**	0.47	0.00	0.00	0.00	0.00	0.00		**0.47**
0.00	0.00	0.00	**0.62, 15.74**	0.61, 14.70	0.00	0.00	0.00	0.00	0.00		**0.61, 14.70**
0.00	0.00	0.00	**0.09**	0.08	0.02	0.01	0.00	0.00	0.00		**0.10**
0.00	0.00	0.00	**0.01, 0.03**	0.02, 0.04	0.00	0.00	0.00	0.00	0.00		**0.02, 0.04**
0.13	0.13	0.00	**3.48**	2.53	0.19	0.31	0.24	0.06	0.00		**3.32**
0.43	0.00	0.00	**6.50**	5.39	0.44	0.33	0.30	0.00	0.00		**6.47**
3.70	4.29	2.40	**50.58**	30.35	4.79	4.39	4.57	3.63	4.15		**51.87**

Costs over 12 and 48 months

We estimated participants' costs in each time period and resource category by multiplying individual resource use in *Table 40* by corresponding unit costs in *Table 38*. We summed estimates across categories to yield the cost of each period during which participants remained in the trial, and discounted costs in later periods at 3.5% per annum (summarised in *Appendix 7.2*). We imputed missing costs (as described in *Chapter 2*, *Statistical methods*, *Imputation and missing values*), and summed participants' costs over 48 months. *Table 41* summarises these costs for the primary cost-effectiveness analysis, starting with costs for the first 12 months. The mean cost of secondary care including hospital prescribed drugs was £2860 lower for the EUS group than for the non-EUS group, even after including the cost of EUS scans. This difference is 9% of the mean cost over 48 months in the control group. In the first year of the trial the difference of £3432 (14% of the control group's mean cost) was even bigger. Both analyses used bootstrapping with 5000 replicates to estimate a 95% confidence interval around this mean difference in costs between groups.

Quality-adjusted life-years

We used survival and EQ-5D data to estimate QALYs to the end of the trial; and imputed from these data to estimate survival and QALYs to 48 months (see *Chapter 2*, *Statistical methods*, *Imputation and missing values*). The following cost-effectiveness analysis uses fully imputed QALYs, summarised in *Table 41*.

TABLE 41 Fully imputed costs and QALYs by allocated group over 12 and 48 months[a]

Type of cost	EUS (*n* = 107): mean (SD)	Non-EUS (*n* = 106): mean (SD)	EUS minus non-EUS (two bootstrapped 95% CIs[b])
Over 12 months			
Secondary care (£)	20,774 (12,346)	23,399 (15,186)	−2624 (−6357 to 1050) or −2624 (−6425 to 1161)
Hospital drugs (£)[c]	994 (1477)	1803 (3976)	−808 (−1744 to −126) or −808 (−1732 to −123)
Total cost/patient (£)	21,769 (13,305)	25,201 (17,953)	−3432 (−7763 to 789) or −3432 (−7829 to 704)
Undiscounted QALYs[d,e]	0.5785 (0.2598)	0.5481 (0.2771)	0.0304 (−0.0416 to 0.1011) or 0.0304 (−0.0441 to 0.1014)
Over 48 months			
Total cost/patient (£)	29,190 (14,902)	32,049 (22,019)	−2860 (−7987 to 2192) or −2860 (−7940 to 2153)
Discounted QALYs (years)[e,f]	1.3616 (0.9989)	1.1647 (0.9756)	0.1969 (−0.0640 to 0.4575) or 0.1969 (−0.0763, to 0.4626)

SD, standard deviation.

a *Appendix 7.2* elaborates on these costs from a different perspective.

b To illustrate variability of this estimate.

c All drugs prescribed in hospital except for chemotherapy or surgery (under secondary care).

d Identical to first row of *Table 27*, because this table does not discount effects in the first year.

e These bootstrapped confidence intervals are very similar to those based on *t*-test in *Table 27*.

f Similar to second row of *Table 27*, which did not discount QALYs and reported a difference of 0.2086.

Cost-effectiveness at 12 months

Primary analysis

By 12 months, 107 participants in the EUS group had on average gained 0.0304 more QALYs and cost £3432 less than 106 participants in the non-EUS group. This suggests that, over the 12 months from randomisation, EUS is more effective and less costly than usual management. However neither difference is significant on its own (see *Table 41*). To enhance the statistical power of such evaluations, cost-effectiveness analysis focuses on the joint distribution of costs and effects. When the intervention under evaluation yields clinical benefits but costs more, the ICER compares these benefits and costs by assessing whether the NHS gains from this trade-off. When the intervention yields both clinical benefits and resource savings, both Briggs *et al*.[132] and Glick *et al*.[131] warn that the resulting negative ICER is meaningless.

Instead we used bootstrapping with 5000 replicates to generate a cost-effectiveness plane (CE plane) plotting the joint distribution of costs and benefits (*Figure 22a*): 3799 (76.0%) points lie in the south-east ('win–win') quadrant, 911 (18.2%) in the south-west quadrant (net costs and QALYs negative), 187 (3.7%) in the north-east quadrant (net QALYs and costs positive) and 103 (2.1%) in the north-west ('lose–lose') quadrant. To summarise the CE plane, we introduce the 'cost-effectiveness threshold' – the most that decision-makers are willing to pay to gain one QALY across patients. By varying this threshold, and thus the proportion of points that achieve cost-effectiveness, we construct the CEAC shown in *Figure 22b*. This shows that: even if decision-makers put no value on an extra QALY, EUS has 94.2% probability (the proportion of points below the *x*-axis in *Figure 22a*) of being cost-effective; when the threshold reaches the lower NICE criterion of £20,000, this probability rises to 95.0%; when the threshold reaches the upper NICE criterion of £30,000, it reaches 95.2%; and, if decision-makers put infinite value on a QALY, it falls to 79.7% (the proportion of points to the right of the *y*-axis in *Figure 22a*). To assess the stability of these bootstrapped probability estimates, we ran another replication; this confirmed that EUS is likely to be cost-effective at 12 months, yielding very similar probabilities – 95.1% at £20,000 and 95.0% at £30,000.

Sensitivity analysis

This conclusion that EUS is cost-effective in managing gastro-oesophageal cancer arises from a small gain in QALYs and large cost savings, mainly because participants in the EUS group spent fewer days as inpatients. However the saving also depends on the unit cost of an EUS scan; if high enough, it could balance cost savings from other sources.

The Department of Health offers a choice of three unit costs: £551 for day cases, £1477 for outpatients and £3781 for inpatients.[152] Our primary cost-effectiveness analysis used the day-case cost of £551 for two reasons: first most trial participants received their scans as day cases; second our own confirmatory costing, based on detailed analysis of the staffing and time needed to deliver EUS to trial participants as day cases, was close to £500.

Figures 22c and *22d* use the unit cost of £1477 for outpatients, thus reducing the mean saving at 12 months to £2628, after adjusting for a few in the EUS group of this pragmatic trial who did not receive their allocated scan and a few in the non-EUS group who nevertheless received a scan. The probability that EUS is cost-effective falls to 89.2% at a threshold of zero, 91.2% at £20,000 and 91.4% at £30,000, but remains stable at infinity. *Figures 22e* and *22f* use the unit cost of £3781 for inpatients, thus reducing the mean saving at 12 months to £626. The probability that EUS is cost-effective falls to 62.0% at zero, 70.8% at £20,000 and 73.5% at £30,000, but again remains stable at infinity. Hence as the unit cost of EUS increases, the probability that EUS is cost-effective at 12 months decreases. As the threshold increases to infinity, however, the probabilities that EUS is cost-effective converge on 79.7% (see *Figure 22b*), 79.3% (see *Figure 22d*) and 80.4% (see *Figure 22f*) respectively, all estimates of the probability that a scan increases QALYs in the first year. The variation in these estimates is inherent in bootstrapping and suggests that our choice of 5000 replicates was appropriate. Thus sensitivity analysis confirms that EUS is probably

(a) Cost EUS = £551 (day case): CE plane
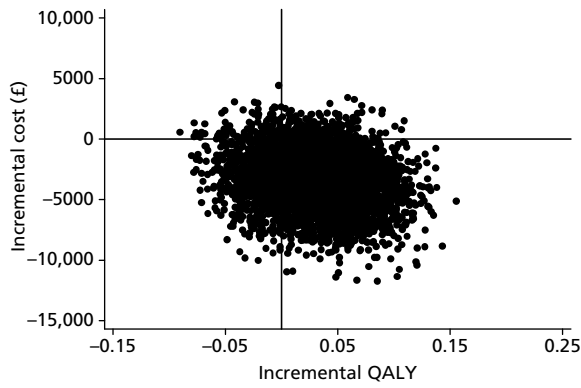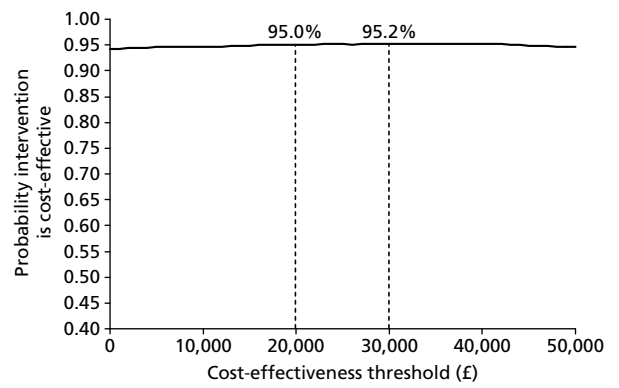
(b) Cost EUS = £551 (day case): CEAC curve

(c) Cost EUS = £1477 (outpatient): CE plane

(d) Cost EUS = £1477 (outpatient): CEAC curve
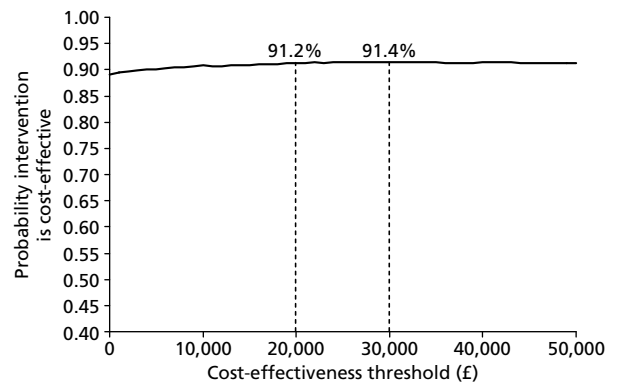
(e) Cost EUS = £3781 (inpatient): CE plane
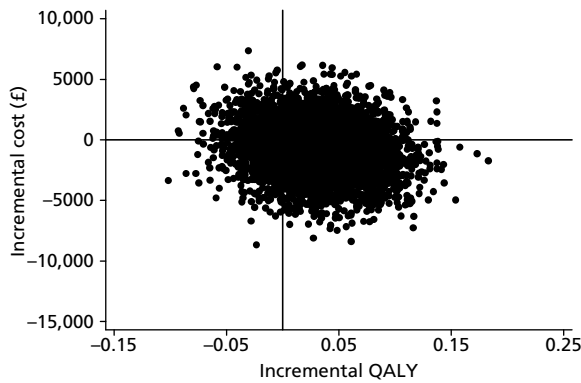
(f) Cost EUS = £3781 (inpatient): CEAC curve
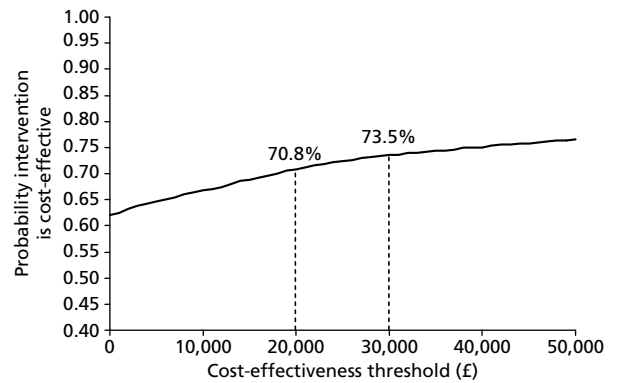
FIGURE 22 Cost-effectiveness plane and CEAC at 12 months by unit cost of EUS.

cost-effective at 12 months; for NICE's preferred thresholds in the range from £20,000 to £30,000 per QALY, the probability that EUS is cost-effective ranges from 71% when a scan costs nearly £4000 to over 95% at a cost of just over £500.

Cost-effectiveness at 48 months

Primary analysis

To extend cost-effectiveness analysis from 12 to 48 months, we kept the unit cost of EUS for day cases at £551, discounted both costs and QALYs at 3.5% per year, and used survival analysis to model likely outcomes for participants who were still alive at the end of data collection (*Figure 23*). Thus by 48 months the 107 participants in the EUS group would on average have gained 0.1969 more QALYs and cost £2860 less than the 106 participants in the non-EUS group. This suggests that, 48 months after randomisation, EUS is still more effective and less costly than usual management.

Again, neither finding is significant. So again we bootstrapped 5000 replicates to generate CE planes and corresponding CEACs. *Figure 23b* shows that the probability of EUS being cost-effective is 86.3% at a threshold of zero; 96.6% at £20,000; 96.8% at £30,000; and 92.6% when the threshold is infinite. To assess stability, we ran another 5000 replicates; these confirmed that EUS is probably cost-effective at 48 months, with very similar probabilities: 97.0% at £20,000 and 96.7% at £30,000.

Sensitivity analysis

As at 12 months we based sensitivity analysis on the alternative unit costs offered by the Department of Health, notably £1477 for outpatients (reducing the mean saving at 48 months to £2055) and £3781 for inpatients (reducing the saving to £53). *Figure 23d* shows that the higher unit cost of £1477 slightly reduces the probability that EUS is cost-effective – to 78.7% at a threshold of zero, 94.8% at £20,000 and 95.3% at £30,000, but leaves it stable at infinity. *Figure 23f* shows that the highest unit cost of £3781 further reduces the probability that EUS is cost-effective – to 48.5% at a threshold of zero, 85.9% at £20,000 and 89.3% at £30,000, but again leaves it stable at infinity. As the threshold increases to infinity, all three probabilities converge on about 92.5%, higher than the corresponding probability at 12 months because the scan yields extra QALYs beyond 12 months. Thus sensitivity analysis confirms that EUS is probably cost-effective at 48 months; for NICE's range of thresholds of cost per QALY, the probability that EUS is cost-effective ranges from 86% to 97%.

Sensitivity analysis at 48 months with undiscounted quality-adjusted life-years

For consistency with undiscounted effectiveness analyses above (see *Effectiveness*, *Quality-adjusted survival: sensitivity analyses with fully imputed data*), we undertook further sensitivity analysis at 48 months, consistent with *Figure 23* in discounting costs at 3.5%, but without discounting QALYs. This increased the average QALY gain for patients in the EUS group a little from 0.1969 to 0.2086, identical to that estimated in *Table 27*. *Figure 24b* shows that not discounting hardly changes the probability of cost-effectiveness, which is: 87.2% at a threshold of zero; 97.0% at £20,000; 96.9% at £30,000; and 92.6% at infinity. As before, we varied unit costs for EUS among those offered by the Department of Health. *Figure 24d* shows that a unit cost of £1477 rather than £551 slightly reduces the probability that EUS is cost-effective – to 79.1% at a threshold of zero, 95.6% at £20,000 and 96.2% at £30,000. *Figure 24f* shows that a unit cost of £3781 further reduces the probability that EUS is cost-effective – to 50.1% at a threshold of zero, 87.2% at £20,000 and 90.1% at £30,000. However the standard deviation (SD) is larger in undiscounted QALYs than in discounted QALYs, so as the threshold increases to infinity, this counteracts the slightly larger QALY gain. Thus all three probabilities converge on the same 92.5% probability of cost-effectiveness as in the primary analysis. In the NICE threshold range of £20,000 to £30,000, each of the three CEAC curves is marginally above the corresponding curve in the primary analysis. Hence all features of *Figure 24* are very similar to those of *Figure 23*.

(a) Cost EUS = £551 (day case): CE plane
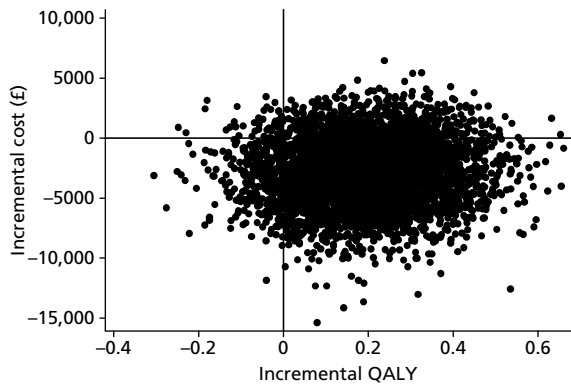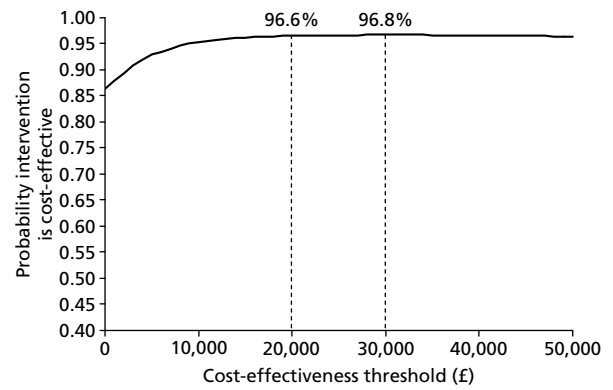
(b) Cost EUS = £551 (day case): CEAC curve

(c) Cost EUS = £1477 (outpatient): CE plane
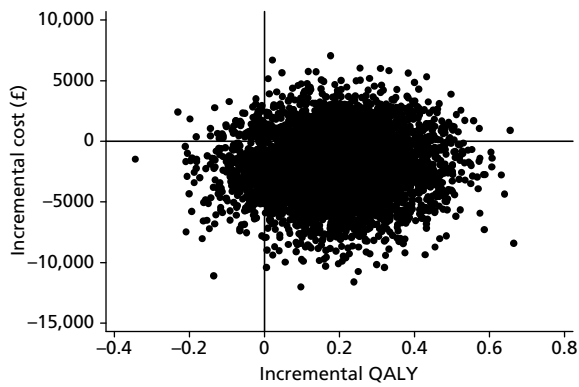
(d) Cost EUS = £1477 (outpatient): CEAC curve
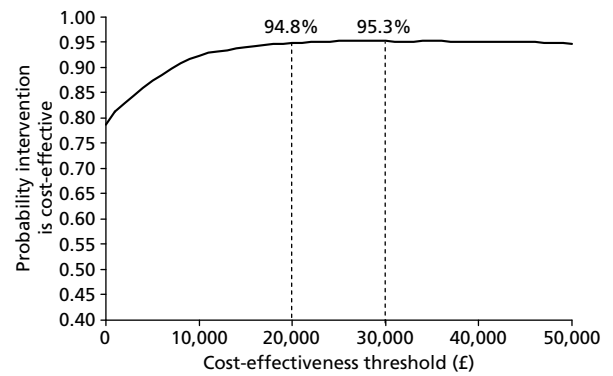
(e) Cost EUS = £3781 (inpatient): CE plane
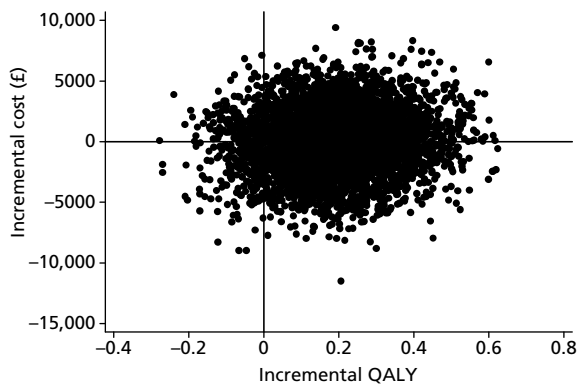
(f) Cost EUS = £3781 (inpatient): CEAC curve
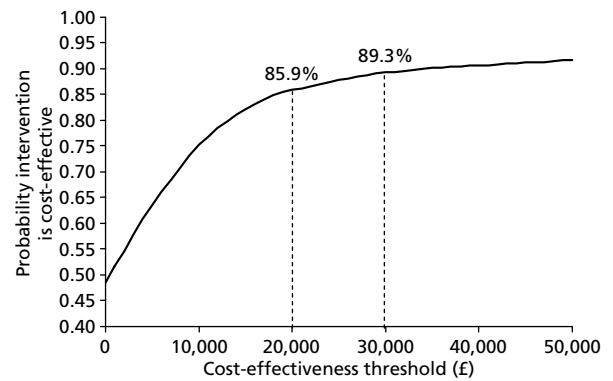
FIGURE 23 Cost-effectiveness plane and CEAC at 48 months by unit cost of EUS.
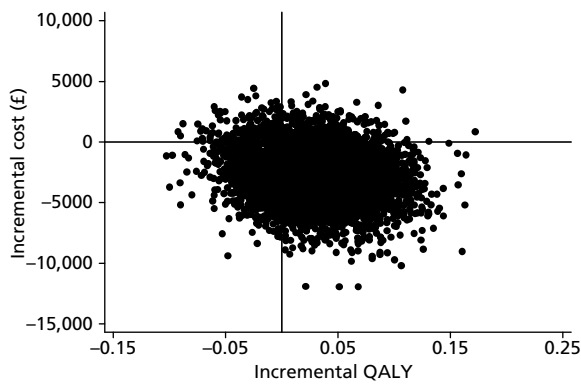
(a) Cost EUS = £551 (day case): CE plane
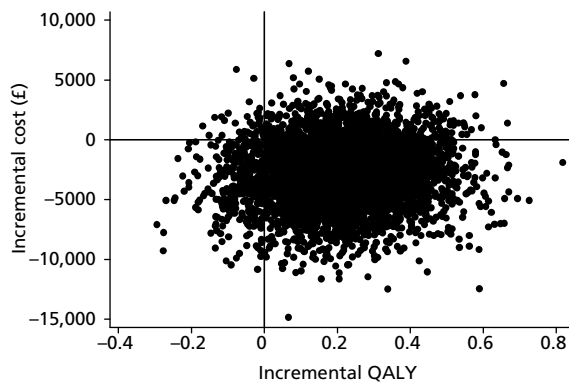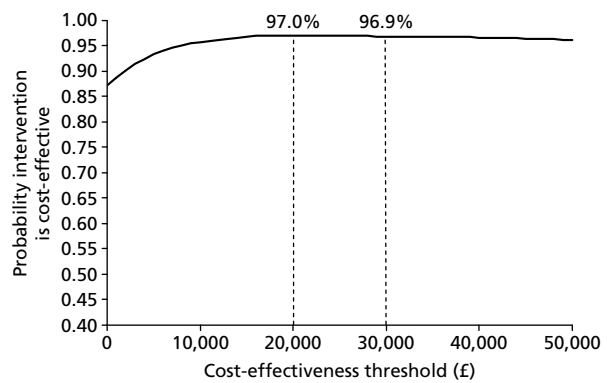
(b) Cost EUS = £551 (day case): CEAC curve

(c) Cost EUS = £1477 (outpatient): CE plane
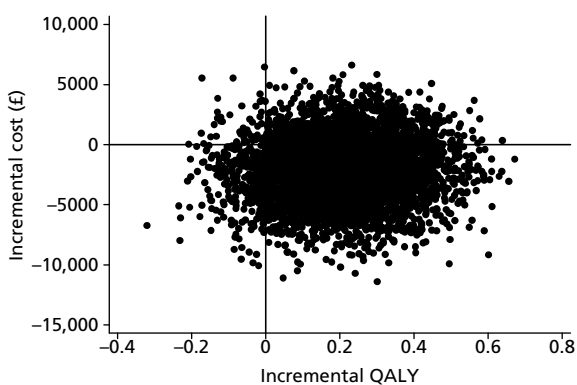
(d) Cost EUS = £1477 (outpatient): CEAC curve
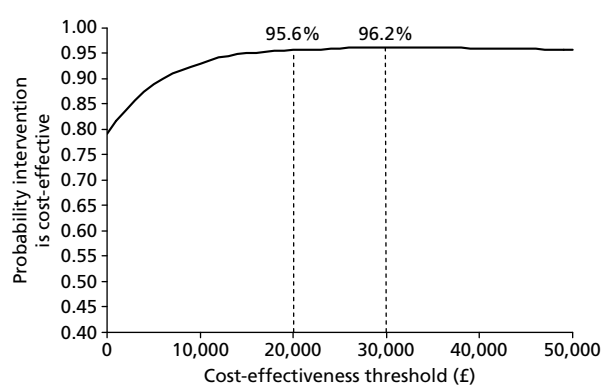
(e) Cost EUS = £3781 (inpatient): CE plane
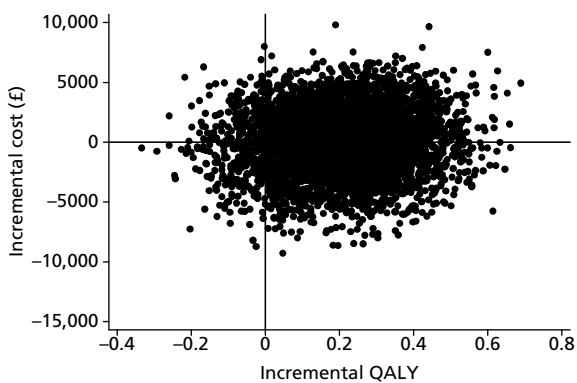
(f) Cost EUS = £3781 (inpatient): CEAC curve
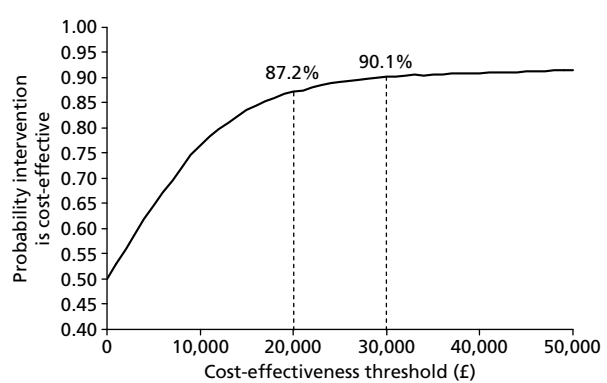
FIGURE 24 Cost-effectiveness plane and CEAC at 48 months by unit cost for undiscounted QALYs.

Subgroup analysis at 48 months

As *Figures 22b* and *23b* provide strong evidence that EUS is likely to be cost-effective in gastro-oesophageal cancer, we ask whether some participants benefit more than others. Our effectiveness analysis addressed this question by adding covariates to primary analysis (see *Table 24*). However cost-effectiveness analysis with bootstrapping is less flexible. Although subgroup analysis reduces sample sizes and statistical power, we examined subgroups defined by baseline EQ-5D (*Figure 25*) and age (*Figure 26*) within the primary analysis to 48 months, in which we discounted both costs and QALYs.

Subgroup analysis by baseline EQ-5D

To maximise statistical power, we split the effective trial population of 213 at the median baseline EQ-5D utility score of 0.796; this divided the EUS group into 54 above that median and 53 below, and the non-EUS group into 47 above and 59 below. Sicker participants had a much higher mean QALY gain (0.3067) than healthier participants (0.0183) but a smaller cost saving (£1918 vs £4259). Combining these effectiveness and cost findings by bootstrapping, we see that EUS in sicker participants has 97.1% probability of being cost-effective at a threshold of £20,000 per QALY, 97.9% at £30,000 and 97.9% at infinity whereas in healthier participants these probabilities fall respectively to 76.6%, 72.5% and 53.3% (still greater than 50% because EUS is more effective and less costly than control, even for high-baseline EQ-5Ds) (*Figure 25b*).

(a) Cost-effectiveness plane
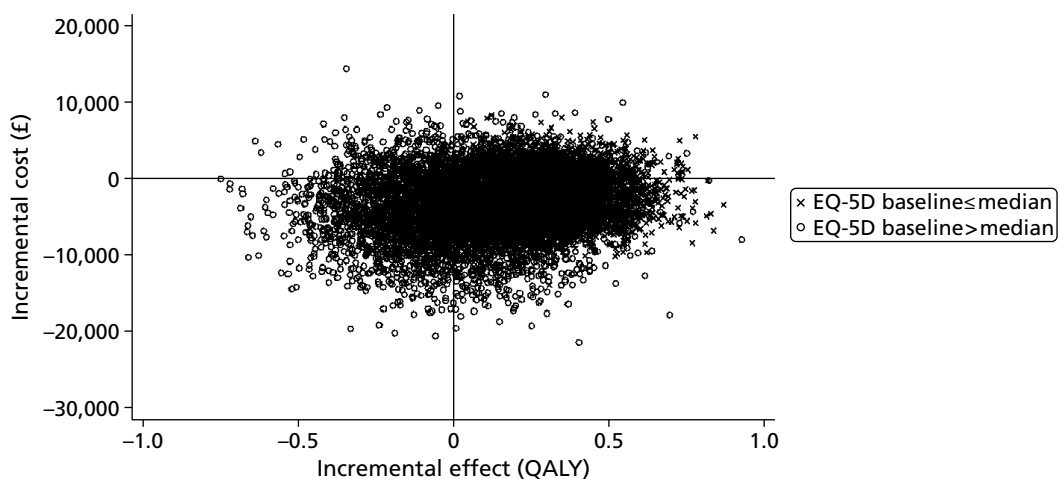
(b) CEAC curve
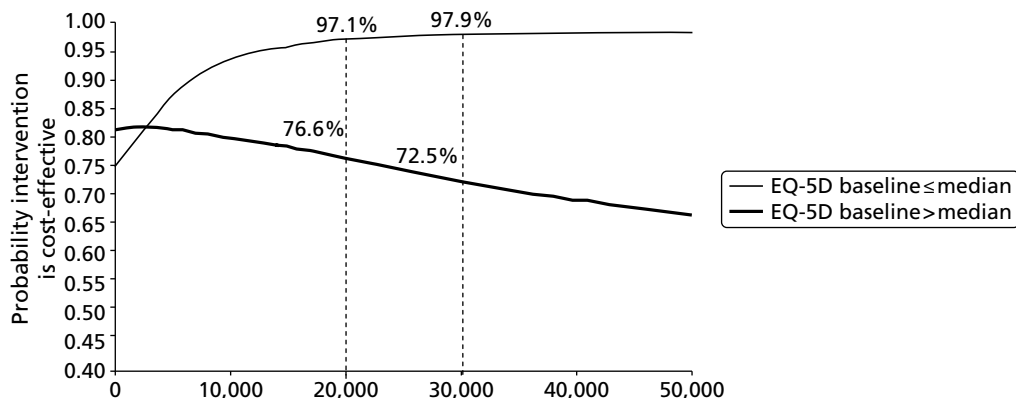
FIGURE 25 Cost-effectiveness plane and CEAC by whether baseline EQ-5D below or above median.

(a) Cost-effectiveness plane

(b) CEAC curve
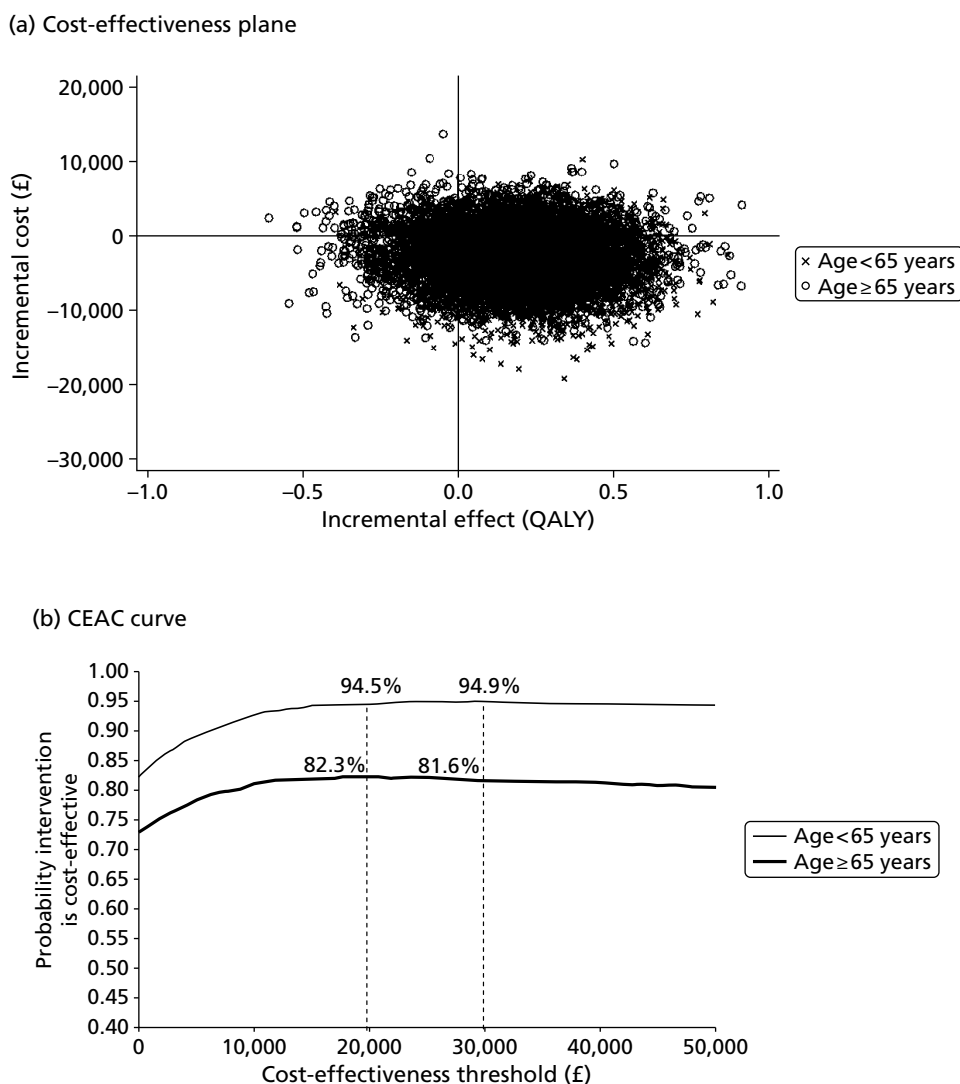
FIGURE 26 Cost-effectiveness plane and CEAC by whether age below or above median of 65 years.

Thus in *Figure 25b* the probability that EUS is cost-effective for healthier participants consistently declines from a peak near £150 with increasing willingness to pay for an additional QALY. The same happens, although less noticeably, in other CEACs – for example at about £40,000 in *Figure 23b* and about £20,000 in *Figure 24b*. Although this shape of CEAC is unusual, it often happens when bootstrapped points cover all four quadrants of the cost-effectiveness plane with cost or effect differences close to zero. In principle the height of a CEAC at a threshold of £0 is equal to the probability that the intervention costs less (i.e. the proportion of points in the south-west and south-east quadrants), rises to a maximum and then declines to the probability that the intervention is more effective (i.e. the proportion of points in the north-east and south-east quadrants). This rise and fall is due to the way in which the threshold changes the balance between the two trade-off quadrants – north-east (where the intervention yields more QALYs but costs more) and south-west (where the intervention yields fewer QALYs but costs less). Although each subgroup analysis uses only about 100 participants, their consistency with the finding that endoscopic ultrasound is significantly more effective for sicker patients (see *Table 24*) strengthens the conclusion that it is also more cost-effective for sicker patients.

Subgroup analysis by age

We split the trial population at 65 years, close to the median; this divided the EUS group into 50 participants of over 65 years and 57 below, and the non-EUS group into 50 over 65 years and 56 below. Younger participants had a higher mean QALY gain (0.2350) than older partipants (0.1528) and a larger mean cost saving (£3454 vs £2246). Bootstrapping shows that EUS in younger participants has 94.5% probability of being cost-effective at a threshold of £20,000 per QALY, 94.9% at £30,000 and 90.6% at infinity; in older participants these probabilities are 82.3%, 81.6% and 76.1% (*Figure 26b*). However this weak analysis provides no evidence that these differences are statistically significant.

Comparing these two analyses, we see that the CEACs in *Figure 25* are farther apart than those in *Figure 26* over thresholds between £20,000 and £30,000. Even with a reduced sample size, the probability of cost-effectiveness between £20,000 and £30,000 is much higher in sicker patients than in the whole sample. To explain why the differential effect of sickness does not translate into a differential effect of age, we observe that the proportions of 'sicker patients' among those under and over 65 years were surprisingly similar – 52% and 53% respectively. In short baseline health status appears to be a better predictor than age of whether EUS is beneficial in gastro-oesophageal cancer.

Summary

Table 42 summarises our health economic findings. All analyses took account of censoring and used bootstrapping with 5000 replicates. Analysis at 12 months after randomisation, covering the period when most treatment occurred, suggests EUS had saved costs (but not quite significantly – see *Table 41*) and gained QALYs (but not significantly – see *Table 41*). Combining these findings we conclude that EUS is significantly cost-effective, in the sense that the probability of being cost-effective exceeds 95% at the NICE threshold of £20,000 to £30,000 per QALY when the cost of a scan is set at the national unit cost of £551 for day cases (*Figure 22b*). However this probability falls below 95% at the less plausible national unit cost of £1477 for outpatient scans (*Figure 22d*) and markedly below 95% at the much less plausible national unit cost of £3781 for inpatient scans (*Figure 22f*). We judge that a cost close to £500 is much more plausible because most trial participants received their scans as day cases and, more importantly, our own detailed analysis of the staffing and time needed to deliver EUS to trial participants as day cases was close to £500.

Analysis at 48 months after randomisation, covering our entire period of useful data collection, with costs discounted at 3.5%, suggests that EUS had saved costs (but not significantly – see *Table 41*) and gained QALYs (but not quite significantly – see *Table 41*). Combining these findings we conclude that EUS is significantly cost-effective at the two lower national unit costs – both £551 for day-case scans and £1477 for outpatient scans. However the probability of cost-effectiveness falls below 95% at the highest national unit cost of £3781 for inpatient scans. Whether or not one discounts QALYs does not affect these conclusions.

Table 24, using Cox regression, shows that participants reporting poorer health at baseline gained significantly more QALYs over a range of follow-up periods averaging 24 months than those reporting better health at baseline. We found a similar pattern, although less significant, in fully imputed 48-month QALYs (see *Effectiveness*, *Quality-adjusted survival: sensitivity analyses for fully imputed data*, and *Appendix 6.18*). Exploratory subgroup analysis at 48 months suggests that these sicker participants saved fewer costs than the healthier participants (but far from significantly – see *Figure 25b*). Combining these findings suggests that EUS is probably more cost-effective for sicker patients. Exploratory subgroup analysis also hints that EUS could be slightly better for younger patients. However these weak subgroup analyses do not detract from the unequivocal finding that EUS is almost certainly cost-effective for gastro-oesophageal cancer.

TABLE 42 Summary of health economics results

No.	Analyses	Discounting	Mean saving (£)	Mean benefit (QALY)	Probability of cost-effectiveness (%)	
					At £20,000	At £30,000
Analyses with EUS unit cost = £551[a]						
1.	Primary analysis at 12 months (costs and benefits undiscounted)	✗ Costs ✗ Benefits	3432	0.0304	95.0	95.2
2.	**Primary analysis at 48 months**	✓ Costs ✓ Benefits	**2860**	**0.1969**	**96.6**	**96.8**
3.	Sensitivity analysis at 48 months (costs discounted but not benefits)	✓ Costs ✗ Benefits	2860	0.2086	97.0	96.9
4A.	Subgroup analysis at 48 months (EQ-5D baseline ≤ median)	✓ Costs ✓ Benefits	1918	0.3067	97.1[b]	97.9[b]
4B.	Subgroup analysis at 48 months (EQ-5D baseline > median)	✓ Costs ✓ Benefits	4259	0.0183	76.6[b]	72.5[b]
5A.	Subgroup analysis at 48 months (age < 65 years)	✓ Costs ✓ Benefits	3454	0.2350	94.5[b]	94.9[b]
5B.	Subgroup analysis at 48 months (age ≥ 65 years)	✓ Costs ✓ Benefits	2246	0.1528	82.3[b]	81.6[b]
Sensitivity analyses at 12 months[a]						
1.[c]	**EUS cost = £551 for day cases (costs and benefits undiscounted)**	✗ Costs ✗ Benefits	**3432**	**0.0304**	**95.0**	**95.2**
6.	EUS cost = £1477 for outpatients (costs and benefits undiscounted)	✗ Costs ✗ Benefits	2628	0.0304	91.2	91.4
7.	EUS cost = £3781 for inpatients (costs and benefits undiscounted)	✗ Costs ✗ Benefits	626	0.0304	70.8	73.5
Sensitivity analyses at 48 months[a]						
2.[c]	**EUS cost = £551 for day cases (costs and benefits discounted)**	✓ Costs ✓ Benefits	**2860**	**0.1969**	**96.6**	**96.8**
3.[c]	EUS cost = £551 for day cases (costs discounted but not benefits)	✓ Costs ✗ Benefits	2860	0.2086	97.0	96.9
8.	EUS cost = £1477 for outpatients (costs and benefits discounted)	✓ Costs ✓ Benefits	2055	0.1969	94.8	95.3
9.	EUS cost = £1477 for outpatients (costs discounted but not benefits)	✓ Costs ✗ Benefits	2055	0.2086	95.6	96.2
10.	EUS cost = £3781 for inpatients (costs and benefits discounted)	✓ Costs ✓ Benefits	53	0.1969	85.9	89.3
11.	EUS cost = £3781 for inpatients (costs discounted but not benefits)	✓ Costs ✗ Benefits	53	0.2086	87.2	90.1

a All analyses took account of censoring and used bootstrapping with 5000 replicates.

b Subgroup analyses with sample size ~ 100 with result that 'probabilities of cost-effectiveness' are intrinsically lower than in full analyses with sample size ~ 200.

c Shaded rows repeated to contribute to two separate comparisons: each shade represents a single pair.

Quality assurance

Process

We requested images from all seven sites which performed EUS scans within the trial and received images from five. Initially only two sites could record scans and forward them by file transfer protocol. Thereafter the trial agreed to buy recording equipment for any site that recruited ten patients into the trial. Variable recruitment meant that only one more site took up this offer. To extend quality assurance to more sites, we obtained still images from one site without recording equipment and, to maintain balance, another site that provided videos.

Figure 27 shows that we assessed 18 videos and two sets of stills (21% of scans performed in the study). Our quality assurance panel held five consensus conferences over the course of the trial, all web-based and audio-facilitated. Contributors included between two and four members but not always those who had provided pre-conference assessments. To maintain objectivity they remained blind to the source of the scans and other assessments until they had reached consensus.
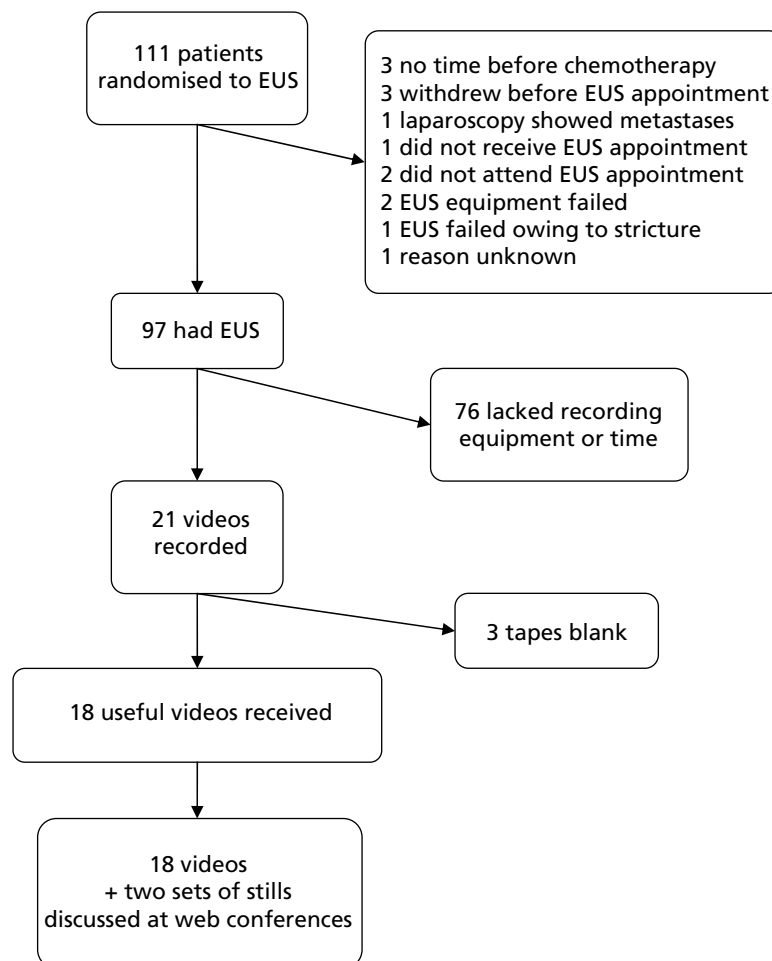
FIGURE 27 CONSORT 'flow chart' of EUS scans for web conferences.

Findings

Images varied in quality, and were generally less clear than an operator would see. The panel judged them adequate for staging, stills less so than videos. Over the five conferences they refined the pre-conference assessment form (see *Appendix 8.2*), for example by adding information about margins. On finding that members differed slightly in classifying borderline nodes, they agreed criteria for nodal involvement: diameter of 1 cm suggests a positive node, as does hyperechoic state. Where the coeliac axis was visible in images, the panel asked assessors to focus on coeliac nodes.

Table 43 displays original assessment, individual pre-conference assessments and blinded consensus. Before the conferences only 6 of the 20 scans showed general agreement between assessors across all stages. Nevertheless the panel reached consensus for all but one T stage and one N stage, including that two T stages and two N stages were inconclusive. For T stage there was a range of individual findings between T1 and T4 but none that varied by more than one from original or consensus; three scans individually graded as T4 became T3 in consensus. As M stage may depend on information outside the scan, both individuals and consensus often graded it as unknown.

Given the consensus that four scans were equivocal, there was no real disagreement between original staging and the final unblinded consensus for 19 T stages and 17 N stages out of 20 tumours (see *Table 43*, footnote 'b'). Even before unblinding, the consensus agreed with the original operator about all T3 classifications, but there was some uncertainty about the boundary between T1 and T2, notably in scans K and R, the only two gastric tumours assessed by the panel. The panel could not see any tumour or node in K but excision revealed a T1 tumour. They could see an involved node but no tumour in R and S. In S, the only tumour originally scored as T1, biopsy had shown a tumour, and the operator had commented 'difficult to see on EUS – slight thickening at 36 cm'; thus the panel understood the original minimum T stage but could not themselves find the tumour. There was occasional disagreement about nodal status: the panel changed Q from N1 to N0, S from N0 to N1, and remained confident that D was N1 even after seeing the original NX. *Table 44* tabulates agreement on T stage between original and blinded consensus, and *Table 45* does the same for N stage.

Thus there was excellent agreement between original observer and panel on T stage (weighted kappa = 0.866; $p < 0.001$), and moderate agreement on N stage (kappa = 0.562; $p = 0.012$). We also asked the panel whether or not each individual scan was a contraindication to resection. They judged that the nine scans that they agreed to stage as T3 N1, all but one in complete agreement with the original staging, made a strong case against resection.

TABLE 43 T, N and M stages by assessor

Image	Original	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6	Consensus
A	T3 N1 M1A	T3 N1 M1A				TX NX MX		T3 N1 M1A
B	T3 N1 M0	T3 N1 M1A	T3 N1 M0			T3 N1 MX		T3 N1 MX
C	T3 N1 M1A	T3 N1 M1A						T3 N1 M1A
D	T3? NX MX	T3 N1 M0	T3 N1 M0					T3 N1[b] M0
E	T3 N0 M0	T3 N0 MX	T2/3 N0 M0		T3 N0 M0			T3 N0 M0
F	T3 N1 M0	T3 N0 MX	TX NX MX		T4 N0 M0			T3 N1 M0
G	T3 N1 M1A	T3 N1 MX	T3 N1 M0		T2/3 N1 MX			T3 N1 M0
H	T3 N0 M0	T3 N0 MX	T3/4 N1 M1		TX N0 MX			T3 N0 M0
I	T2 N1 MX	T3 N1 MX		T2 N1 M1	T2 N1 M0?	T1(2?) N1 M–	T3 N1 MX	T2 N1 MX
J	T2 N0 M–	T3 N1 MX		T1 N0 MX	T1 N0 M0	T1 N0 M–	T3/4 N1 M–	T2 N0/1 [c,d] MX
K	T2 N0 M0	T3 N0 M–		TX NX M–	T1 N0 M0	T1 N0 M–	T2 N0 M0	T0[b] N0 M0
L (still)	T3 N0 M1A	T3 N1 M1A		T3/4 N1 M–	T3 N0? MX	T2(3?) N1 M–	TX NX MX	T3 N0/1[c] M1A
M	T2 N0 M0	T2 N0 M0		T2 N0 MX	T1 N0 MX	T1 N0 MX		T2 N0 M–
N	T3 N1 M–	T3 N1 M0		T3 N1 MX	T4 N1 MX	T3 N1 MX		T3 N1 MX
O	T2 N1 M0	T2 N1 M0		T1 N0/1 MX	T1 N0 MX	T1 N0 MX		T1/2 [c,d] N0/1[c] MX
P (still)	T3 N1 M–	T3 N0 M0		TX NX MX	T3 N1 MX	T3 N1 MX		T3 N1 M0
Q	T2 N1 M0	T2 N1 M0/1A		T2 N1 MX	T3 N0 MX	T3 N1 MX		T2 N0[b] MX
R	T2 N1 M0	TX NX MX		TX N1 M–	TX N1? MX	T2 N0 MX		T1/2[c] N1 M–
S	T1 N0 M0	T2 N1 M0/1A		TX N1 M1	T1 N1 MX	T1 N0 MX		T1/2[c] N1[b] M–
T	T3 N1 M0	T2 N1 M0/1A		T3 N1 MX	T3 N1 MX	T4 N0 MX		T3 N1 M0

X, not sure; –, no response.
a Panel differed from original, whether panel blinded to original or not (D, K, Q and S for N).
b Panel differed from original while panel blind to original, but not once unblinded (J, L, O, R and S for T).
c Panel could not achieve consensus (J and O).

TABLE 44 T stage: agreement between original operator and consensus panel

| Original | Blinded consensus | | | |
	T0 or T1	T2	T3	Total
T1	0.5	0.5	0	1
T2	2	5	0	7
T3	0	0	12	12
Total	2.5	5.5	12	20

Note: The three tumours graded T1/2 by consensus count as 0.5 in T1 and 0.5 in T2.

TABLE 45 N stage: agreement between original operator and consensus panel

| Original | Blinded consensus | | |
	N0	N1	Total
N0	5	2.5	7.5
N1	1.5	11	12.5
Total	6.5	13.5	20

Note: The four tumours graded NX originally or N0/1 by consensus count as 0.5 in N0 and 0.5 in N1.

Table 46 shows that staging by the external assessor, using an adapted video assessment form (see *Appendix 8.3*) and an anonymised summary of the original operator's proforma (see *Appendix 8.1*), differed more from the original operator than the panel did. The external assessor agreed with the T stages of 11 scans, but upstaged seven and was unsure about the other two. Of the 12 originally graded as T3, he agreed with seven but thought that five were T4. Hence he had a lower threshold for scoring pleural involvement. He upstaged two other scans from T2 to T3; and four nodal assessments, three in common with the consensus panel.

Three internal assessors re-examined five videos in the knowledge that the external assessor had scored them as T4 (*Table 47*). One assessor commented 'Pleural involvement does not seem to be a factor in patient management at our MDT meetings although I appreciate it does alter the T stage from T3 to T4.'

Our assessors agreed that T4 was possible in four of these five scans. Once one assessor suggests pleural involvement, perhaps others are more likely to find it than if they viewed the films blind.

Of these five patients, four had chemoradiotherapy and one had neo-adjuvant chemotherapy before surgery. Hence there is no evidence from excised tumours that could confirm or refute the T4 scores. Thus the only corroborative evidence comes from survival. As mean survival was 371 days (median 188, range 141 to 810), it is unlikely that all had T4 cancers.

TABLE 46 T, N and M stages, including external assessor's, with time and cause of death

Image	Staging at randomisation	Original EUS	Days from EUS to death	Cause of death	Consensus	External assessor
A	T3 N1 M0	T3 N1 M1A	810	Cancer	T3 N1 M1A	T4[b] N1 M1A
B	T3 N1 M0	T3 N1 M0	556	Cancer	T3 N1 MX	T4[b] N1 M0
C	T2 N0 M0	T3 N1 M1A	522	Cancer	T3 N1 M1A	T3 N1 M1
D	No record	T3? NX MX	161	Treatment related	T3 N1 M0	T4[b] N1[b] MX
E	T3 N1 M0	T3 N0 M0	–	Alive	T3 N0 M0	T3 N0 MX
F	T3 N1 M0	T3 N1 M0	163	Cancer	T3 N1 M0	T3 N1 M0
G	T3 N1 M0	T3 N1 M1A	584	Cancer	T3 N1 M0	T3 N1 M1A
H	T2 N0 M0	T3 N0 M0	133	Treatment related	T3 N0 M0	T3 N0 MX
I	T2 N1 M0	T2 N1 M?	–	Alive	T2 N1 MX	T3[b] N1 MX
J	T3 N1 M0	T2 N0 M–	–	Alive	T2 N0/1 MX	T2 N0 M–
K	T1 N0 M0	T2 N0 M0	–	Alive	T0 N0 M0	TX[b] NX[b] MX
L[c]	T3 N1 M0	T3 N0 M1A	318	Other (cancer contributed)	T3 N0/1 M1A	T3 NX[b] M1A
M	T3 N0 MX	T2 N0 M0	–	Alive	T2 N0 M–	T2 N0 M0
N	T3 N1 MX	T3 NI M–	–	Alive	T3 N1 MX	TX[b] N1 MX
O	T2 N0 M0	T2 N1 M0	144	Treatment related	T1/2 NX MX	T2 N1 M–
P[c]	T3 N1 M0	T3 NI M–	188	Other (cancer contributed)	T3 N1 M0	T4[b] N1 MX
Q	T2 N0 MX	T2 N1 M0	–	Alive	T2 N0 MX	T3[b] N1 M0
R	T3 N1 M0	T2 N1 M0	–	Alive	T1/2 N1 M–	T2 N1 M0
S	No record	T1 N0 M0	–	Alive	T1/2 N1 M	T1? N1[b] M0
T	No record	T3 N1 M0	141	Other (cancer contributed)	T3 N1 M0	T4[b] N1 MX

X, not sure; –, no response.

a External assessor differed from original in T or N.

b Still image.

TABLE 47 Tumours graded T4 by external assessor: response from internal assessors

Film	Assessor	Response to suggestion of pleural involvement
A	1	Yes, could be pleural involvement
A	4	Suspicious for pleural involvement
A	5	Not pleural, possible pericardial
B	1	Yes, could be pleural involvement
B	4	No definite involvement
B	5	No, but poor images
D	1	Yes, could be pleural involvement
D	4	Suspicious, not definite pleural involvement
D	5	Significant shadowing from calcification, so cannot be sure – I think not
P (still)	1	No pleural involvement
P (still)	4	No pleural involvement
P (still)	5	No pleural involvement
T	1	Yes, could be pleural involvement
T	4	Pleura look involved
T	5	Yes, could be pleural involvement (from 1:58 to 2:09 in video)

Comparison of endoscopic ultrasound staging with pathology of excised tumours

Our secondary outcome measures include 'pathological reporting of resected specimens using SAGOC criteria' (see *Chapter 2*, *Design*, *Secondary outcome measures*). To this end *Table 48* compares staging before randomisation, staging by EUS and excisional staging for the only 21 intervention group participants who had surgery without prior treatment, as they alone can contribute without bias to quality assurance of EUS. As there may have been a gap of a few weeks between staging and resection, the true stage may have advanced. However spontaneous reduction in stage is so unlikely that preoperative overstaging is the more likely explanation for reductions in stage. Although we have pathology reports for tumours resected after chemotherapy or radiotherapy, they are biased by that intermediate treatment and cannot contribute usefully to quality assurance.

Of 21 participants in *Table 48*, four staged as T3 by EUS were T2 on histopathology, and two staged as T2 became T1; three staged as T2 were T3 on histopathology, one staged as T2/3 became T3, and one staged as Tis became T2. As one scan and one pathological examination were incomplete, only eight participants showed agreement on T stage. Of the 19 participants for whom EUS could estimate N stage, seven missed nodal involvement and two over-reported it – not a significant difference; thus only 10 patients showed agreement on N stage.

Again quality assurance needs the three-dimensional *Tables 49* and *50* to focus on the same 21 uncontaminated participants undergoing EUS as *Table 48* and to exclude participants receiving neo-adjuvant therapy, because intermediate treatment biases estimated stage; where T or N stage is missing from early staging, endoscopic staging or excisional staging, numbers fall below 21. *Table 49* tabulates agreement on T stage between these three estimates, and *Table 50* does the same for N stage. There is no statistical evidence here that EUS is a better predictor of histopathology than early staging is.

TABLE 48 Staging for EUS participants who had surgery without pre-treatment

Resection	Surgery	Early staging	EUS staging	Excision staging
Complete	Subtotal gastrectomy + Roux-en-Y	T3 N1 MX	T3 N1 M0	T3 N1 M1
Complete	Gastrectomy: 1/7 nodes involved	T3 N0 M0	T3 N0 M0	T2[a] N1[b] M0
Complete	Oesophago-gastrectomy	T3 N1 M0	T3 N1 M0	T3 N0[a] M0
Complete	Oesophago-gastrectomy: 3/8 nodes involved	T2 N0 M0	T2 N0 M0	T– N1[b] M–
Complete	Total gastrectomy + Roux-en-Y	T4 N0 MO	T2[c] N0 MX	T3 N1[b] M0
Complete	Oesophagectomy + feeding tube	T3 N1 M0	T2 N1 M0	T3[b] N1 M0
–	Subtotal gastrectomy	T1 N0 M0	T2 N0 M0	T1[a] N0 M0
Complete	Partial gastrectomy + Roux-en-Y	T– N– M0	T2 N1 MX	T1[a] N1 M0
Complete	Total gastrectomy + Roux-en-Y	T3 N1 MX	T2 N1 M0	T3[b] N1 MX
Complete	Gastric resection	T3 N1 MX	T3 N0 M0	T3 N1[b] M1
Complete	Oesophago-Gastrectomy	T1 N0 M0	T1 N0 M0	T1 N1[b] M0
–	Subtotal gastrectomy	T2 N0 MX	T2/3 N0 M0	T3b N1[b] MX
Complete	Oesophageal resection	–	T1 N0 M0	T1 N0 MX
Complete	Subtotal gastrectomy	T1 N0 MX	T2 N1 M0	T2 N1 MX
Complete	Gastric resection	T2 N0 M0	T3 N1 M0	T2[a] N0[a] MX
Complete	Gastric resection	T2 N0 M0	T3 N1 M1	T3 N1 MX
Complete	Gastric resection	–	T3 N1 M0	T3 N1 MX
Complete	Gastric resection	–	T3 N– MX	T2[a] N1 MX
Complete	Gastrectomy + feeding tube	–	T2 N0 M–	T3[b] N1[b] MX
Complete	Gastrectomy	Tis N0 M0	Tis N– M–	T2[b] N0 M0
Complete	Oesophagogastrectomy	T3 N1 M0	T3 N1 MX	T2[a] N1 MX

x, not sure; –, no response.
a Likely overstaging by EUS.
b Potential understaging by EUS.
c EUS scan not complete – could not pass tumour.

TABLE 49 Agreement between EUS T stage and excision biopsy reports

Early staging	EUS staging			
	Tis or T1	**T2**	**T3 or T4**	**Total**
Tis or T1	2	2	0	**4**
T2	0	1.5[a]	2.5[a]	**4**
T3 or T4	0	3	5	**8**
Total	**2**	**6.5**	**7.5**	**16**
Early staging	**Excisional staging**			
	Tis or T1	**T2**	**T3 or T4**	**Total**
Tis or T1	2	2	0	**4**
T2	0	1	2	**3**
T3 or T4	0	1	7	**8**
Total	**2**	**4**	**9**	**15**
EUS staging	**Excisional staging**			
	Tis or T1	**T2**	**T3 or T4**	**Total**
Tis or T1	2	1	0	**3**
T2	2	1	4.5[a]	**7.5**
T3 or T4	0	4	5.5[a]	**9.5**
Total	**4**	**6.**	**10**	**20**

a The tumour graded T2/3 by EUS counts as 0.5 in each stage.

TABLE 50 Agreement between EUS N stage and excision biopsy reports

Early staging	EUS staging		
	N0	**N1**	**Total**
N0	6	3	**9**
N1	1	5	**6**
Total	**7**	**8**	**15**
Early staging	**Excisional staging**		
	N0	**N1**	**Total**
N0	3	7	**10**
N1	1	5	**6**
Total	**4**	**12**	**16**
EUS staging	**Excisional staging**		
	N0	**N1**	**Total**
N0	2	7	**9**
N1	2	8	**10**
Total	**4**	**15**	**19**

Chapter 4 Discussion

Endoscopic ultrasound has the potential to stage gastro-oesophageal cancers accurately, and thus to provide prognostic information to guide management. There is widespread evidence about the accuracy of EUS in staging gastro-oesophageal cancer,[35,36] and as a result there have been many recommendations to use EUS for this purpose. However there has been no rigorous evaluation of the effectiveness of EUS in changing clinical management[143] or in improving patients' health and well-being. Hence the supposed link between better staging, better management and better patient outcomes is not yet proven.

So in 2004 the NHS HTA programme commissioned the COGNATE team to conduct a pragmatic randomised trial to assess whether the health technology of EUS, when added to the usual staging tests, changes treatment, improves survival and quality of life, and uses resources cost-effectively. In the resulting study we successfully allocated 223 patients at random between an intervention group, to whom EUS was available to enhance the usual staging algorithm, and a control group limited in principle to that algorithm. Of these trial participants, 213 (96%) yielded enough data for primary analysis.

Summary of findings

Psychometric analyses

From the FACT portfolio of outcome measures we combined existing gastric and oesophageal site-specific modules into a single comprehensive scale. We gave equal weight to the four subscales when combining them for FACT-G. Following internal analysis of FACT scales we dropped two items – one social and one functional – which did not perform well.

The psychometric analyses supported both our choice of EQ-5D as the primary means of adjusting survival in this trial and our decision to supplement it by FACT measures which include aspects of individuals' experience not captured by EQ-5D, namely their social and emotional well-being. Hence our effectiveness analyses used EQ-5D and two FACT summary measures: FACT-G, an average of the four subscales, and FACT-AC, combining the site-specific gastric and oesophageal modules. Structural modelling indicated that the alternative FACT summaries (TOI and FACT Total) were less appropriate than the two selected; while EQ-VAS correlated moderately well with EQ-5D and the physical and functional FACT scales.

Our original plan for analysing each FACT scale considered its own baseline values but not the baselines of the other scales. However the existence of a strong interaction between EUS and baseline EQ-5D in the QALY and 12-month EQ-5D models prompted us to include EQ-5D as an extra covariate in analysing FACT. To reduce the chance of false-positives we did not consider all baseline scale and subscale values as potential covariates.

Although comparisons of FACT between groups are not significant with or without interaction, those for FACT-adjusted survival favour EUS. There are highly significant interactions with baseline EQ-5D but not with the baseline of the corresponding FACT scale or subscale. Even when this is the only interaction in the model, the effects of EUS and baseline FACT-G on FACT-G at 12 months are unrelated. Although each baseline FACT has a considerable effect on the corresponding 12-month follow-up and on FACT-adjusted survival, that effect is present equally in EUS and non-EUS groups, and so does not affect the difference between them. Although the finding that EQ-5D is a better covariate for FACT than FACT itself is unusual, the findings from FACT are entirely consistent with, and serve to reinforce, those from EQ-5D.

Effectiveness

At the end of the trial, 44% of intervention participants, allocated at random to EUS, and 32% of control participants, allocated to usual staging, were alive. EUS improved survival adjusted for generic quality of life (EQ-5D), i.e. QALYs. The resulting hazard ratio was 0.705 (95% CI from 0.499 to 0.995), reflecting a difference in the estimated median quality-adjusted survival between 1.12 years in the intervention group and 0.94 years in the control group – a difference of 66 days in perfect health.

Both components of the composite primary outcome measure – survival and patient-reported HRQoL in the form of EQ-5D scores – also differed between groups. The estimated median survival time was 1.96 years in the intervention group, compared with 1.63 years in the control group – a difference of 121 days, albeit at reduced quality of life. Allowing for covariates would increase that difference; the corresponding hazard ratio was 0.706 (95% CI from 0.501 to 0.996) compared with 0.758 in a simple comparison. The mean participant-reported EQ-5D scores at 12 months were 0.509 in the intervention group and 0.449 in the control group – a non-significant difference of 0.061 (95% CI from –0.043 to 0.164). However the effectiveness of EUS did not differ between centres, cancer sites or cancer stages.

In contrast, the benefits of EUS were markedly greater for participants with poor initial quality of life; in statistical terms there was significant interaction between EUS and generic quality of life (EQ-5D) for all these outcomes. There was also significant interaction between initial generic quality of life and the effect of EUS on all relevant scales within FACT, both at 12 months and in the form of FACT-adjusted survival. Again sicker patients benefitted more from EUS. However there was no interaction between initial FACT scores and the effect of EUS on FACT-based outcomes, and no significant difference between intervention and control groups in mean FACT scores adjusted for covariates.

Both management plans and final treatment varied between centres. Although EUS changed the management plan for several participants, differences between groups in the proportion who received each treatment were not significant. Although the proportion of attempted resections that were incomplete was twice as high in the control group (20%) as in the EUS group (9%), this was not significant because there were only 108 real attempts at resection. Nevertheless the contrasting survival experience of intervention and control participants whose management plans changed is consistent with the hypotheses that: EUS identifies participants whose initial plans were not appropriate; and the resulting change of treatment contributes to improved survival and quality of life.

Cost-effectiveness

For both intervention and control participants we estimated the total cost to the NHS of health care during their time in the trial. We aggregated the costs they incurred in outpatient clinics, as day patients, during inpatient stays and through prescribing. Even allowing for the cost of the EUS scan itself, participants in the intervention group generally consumed fewer resources in secondary care and pharmaceuticals. The average total cost of care over 48 months was £29,200 (SD £14,900) in the intervention group but £32,000 (SD £22,000) in the control group. Hence allocation to EUS saved an average of £2860. The large standard deviations show that these cost data are skewed, as usual. So the best way to estimate confidence intervals is by 'bootstrapping', i.e. by resampling 5000 times from the original distribution of costs. This leads to an estimated confidence interval for the saving due to EUS of between –£2200 and £8000.

Bootstrapping is also useful for combining the data on quality-adjusted survival with those on resource costs. We used the same 5000 sample points to compare the estimated net benefit in QALYs from EUS with the estimated net cost of EUS. Of these, 3988 lie in the south-east quadrant of the cost-effectiveness plane where EUS both increases QALYs and saves NHS costs – best summarised as 'win–win' for EUS; a mere 46 points lie in the north-west quadrant where EUS both loses QALYs and increases NHS costs ('lose–lose'); 640 points lie in the north-east quadrant where EUS improves benefits but increases costs; and 326 in the south-west quadrant where EUS makes savings but loses QALYs.

Since NICE sets the threshold for cost-effectiveness between £20,000 and £30,000 per QALY,[148] bootstrapping leads to the conclusion that the probability that the EUS gives 'value for money' lies between 96.6% and 96.8%. Even if the NHS is unwilling to pay for an additional QALY, the probability that EUS is cost-effective is (3988 + 326)/1000, namely 86.3%. Finally, if the NHS were to set its cost-effectiveness threshold very high, there would still be (3988 + 640)/1000 = 92.6% probability that EUS would be cost-effective.

Quality assurance

This study showed the feasibility of undertaking quality assurance as part of a multicentre randomised trial of diagnostic technology. Five web-based consensus conferences allowed peer review of 20 anonymised scans (21% of 97 scans performed within COGNATE) that would not have been possible face to face because of the physical distances between reviewers. The conferences facilitated discussion, consistently leading to consensus among the panel representing six centres.

The process established that interpretation of scans varied between centres, and that peer review could alter final assessment of T (tumour) and N (nodal) stages. Most critical decisions relate to the diagnosis of N1 or T4 disease, which change management if the tumour has previously been diagnosed as less advanced (namely N0 or T3). Against this background the panel upgraded two scans to N1 and downgraded another to N0. They also reclassified one T stage from T2 to T0; thus it may have been suitable for a less radical resection such as EMR. Changes in M (metastatic) stage from M0 to M1 would also have changed clinical decisions but this did not occur within our web conferences. In short, variation between assessors had little effect on clinical decision-making, in particular whether to resect cancers.

To enhance this quality assurance we recruited an external assessor from another Commonwealth country who was very familiar with EUS scanning. Of the five tumours that the external assessor graded as T4, only one went on to resection. Our panel graded fewer scans as T4. Although two panel members graded three different tumours as T4, the panel later agreed that all three were T3. Hence there is benefit in reviewing scans when the decision is likely to affect clinical decisions.

The panel could not always verify the nodal assessment of the coeliac axis. Yet all but two of our consensus assessments stemmed from full EUS scans. This may highlight a training issue in using EUS in this context. Nevertheless kappa statistics showed a moderate level of agreement between assessors and original investigators on nodal involvement. This is important because the presence of nodal disease was the commonest reason why EUS changed treatment decisions away from cancer resection.

Strengths and weaknesses of the COGNATE trial

General

In 2001 the HTA programme responded to the systematic review by Harris *et al.*[143] by issuing a commissioning brief for a pragmatic randomised trial of EUS as an adjunct to the usual staging tests for gastro-oesophageal cancer. The successful bid came from an alliance between the leaders of SAGOC[4] and a small team of trialists in the emerging North Wales Clinical School. Experience of SAGOC and the sparse literature had convinced them that the case for EUS was unproven.

Unfortunately the commissioning process took longer than usual, for reasons beyond the control of both the HTA co-ordinating centre and the COGNATE team. As a result COGNATE began recruitment 7 years after the systematic review that had triggered it. Not surprisingly most UK centres had by then decided that EUS should be part of routine practice for gastro-oesophageal cancer. The original COGNATE bid in 2002 had included genuine expressions of interest from eight centres. Another eight seriously considered taking part after COGNATE had achieved funding, often by inviting the trial team to present their case to the local MDT. Nevertheless in most centres there were clinicians whose personal experience had led them to believe that EUS was effective and who would not forgo EUS for any patient. However two

centres – Aberdeen and Gloucester – remained in equipoise about the practical value of EUS and recruited participants effectively over 3.5 years. Another six centres allowed individual clinicians in equipoise to recruit participants as and when they could. By analysing these six as one centre, the COGNATE team has in effect delivered the proposed trial in the equivalent of three centres rather than the target of 10. Although the trial team has therefore done its best under difficult circumstances, it was arguably risky to proceed with the COGNATE trial when only two UK centres were truly in equipoise.

In the face of opposition to the proposed trial, the COGNATE team, supported by the TSC and DMEC, maintained its commitment to rigorous evaluation to underpin evidence-based practice. In particular we continued to encourage the remaining centres to recruit whenever they could. We also revised our power calculations in two ways. We responded to the initial realisation that many centres had already adopted EUS by replacing our primary outcome of survival by quality-adjusted survival, often called quality-adjusted life-years, for two reasons: (1) this reduced the target sample size from 700 to 400; and (2) QALYs is the criterion preferred by NICE. We judged that this was the best we could do to generate trial-based evidence in the face of the widespread adoption of EUS. As the number of actively recruiting centres continued to fall, we increased the target effect size for both survival and quality of life from 0.3 to 0.4 (still generally regarded as a 'small' effect). That enabled us to reduce the target sample size to 220. To achieve even this minimal target we had to extend the period of recruitment by 6 months – from 3 to 3.5 years.

Thus the main weakness of the COGNATE trial is that it was some 10 years too late. Committed to the practice of evidence-based medicine, however, we judged that rigorous evaluation was still essential. So we strove in four distinct ways to deliver the best trial we could in difficult circumstances.

First we used information technology in the form of laptop computers and file transfer protocols to facilitate data collection in all eight centres.

Secondly we complied with GCP throughout. Over the life of COGNATE we translated the principles of GCP into a portfolio of practical SOPs, thus laying the foundations of NWORTH, now a Registered Clinical Trials Unit. In particular, the economic analysis of COGNATE follows our SOP for economic evaluation alongside RCTs,[141] based on the authoritative text by Glick *et al.*[129]

Thirdly we successfully included an assiduous representative user within trial management – both TSC and DMEC – at a time when 'patient and public involvement' (sic) was in its infancy.

Finally we pursued the principle of 'analysis by treatment allocated'[112] in the face of a modicum of missing data, notably by following the four steps later advocated by White *et al.*[159]

1. Try to follow up all randomised participants, even if they withdraw from allocated treatment.
   We defined two levels of withdrawal – complete withdrawal and withdrawal from questionnaire interviews. Hence local researchers could stress that withdrawing from allocated treatment or active participation did not require withdrawal from data collection, and continue to collect data from hospital records. Thus only six participants withdrew completely and 25 partial withdrawals contributed to secondary analysis (see *Figure 8*). Rigorous data collection was generally successful in minimising missing data (see *Chapter 2*, *Design*, *Electronic data collection and storage*).
2. Perform main analysis of all observed data under plausible assumption about missing data.
   Loss of data occurred mainly through censoring at the end of the trial, one to five years after randomisation. As it is safe to treat censoring as 'missing at random',[129,131] both survival analysis for effectiveness and survival-based imputation for cost-effectiveness are valid (see *Chapter 2*, *Statistical methods*, *Imputation and missing values*).
3. Perform sensitivity analyses to explore departures from this 'plausible assumption'.
   Our main sensitivity analysis tested the validity of imputing missing data on the assumption that they are 'missing at random'. The first confirmed that our primary 48-month cost-effectiveness analysis was consistent with secondary analysis restricted to the first year of the trial, and with the

primary effectiveness analysis of quality-adjusted survival, both of which used little imputation. The second confirmed that estimated quality of life at later follow-up times yielded consistent findings whether analysed for all participants with imputation beyond the end of the trial where necessary, or else for all (dead or alive) who reached that time-point before the end of the trial (31 July 2009), which used less imputation.

4. Account for all randomised participants, at least in the sensitivity analyses.
    Only two participants (one from each group) withdrew so soon that they could not contribute to any analysis; moreover, both primary analyses – effectiveness and cost-effectiveness – used 213 participants (96% of 223). So sensitivity analysis of the missing participants would thus have been trivial.

Trialists, notably Schwartz and Lellouch[109] and the CONSORT group,[110] have consistently argued that 'analysis by treatment allocated' (previously known as 'analysis by intention to treat') is the only unbiased approach to analysing trials. Nevertheless we undertook 'analysis by treatment received' as another form of sensitivity analysis (see *Table 37*). The 14 participants who did not receive EUS as allocated included three for whom there was no time to arrange appointments. The implication is that surgery was the unequivocal management plan. Hence transferring these three from the EUS (allocated) group to the non-EUS (received) group increases the proportion of changed plans in the former and reduces it in the latter, thus biasing analysis with changes in plan as criterion. Nevertheless the sensitivity analysis showed findings consistent with the primary analysis.

These developments in the conduct of randomised trials enabled the two main clinical centres to achieve high standards of recruitment and data collection and the team in Bangor to create a data set of high quality in a sick population. That COGNATE achieved largely unequivocal findings despite recruiting only a minimal sample is a measure of the rigour of these processes.

Effectiveness and cost-effectiveness

By changing our primary outcome prospectively, we identified significant positive findings, both for survival (the original primary outcome) and for quality-adjusted survival (the eventual primary outcome). How reliable are these findings in the light of the small sample? As the six centres outside Aberdeen and Gloucester contributed only 20% of participants, we evaluated EUS mainly in two UK centres still in equipoise about its role when we began the trial in 2005, perhaps the only two. These two centres remained in equipoise about the practical value of EUS and recruited participants enthusiastically and effectively over 3.5 years. The other six centres allowed individual clinicians in equipoise to recruit participants as and when they could. To get the most out of these 43 participants, we treated them as coming from a virtual third centre.

Nevertheless there were 120 potential participants who could not enter the trial because EUS was considered essential for them (see *Figure 4*). As a result, COGNATE validly randomised only 223 (47%) of 477 patients adjudged eligible to join the trial. So who was left in the trial? What does this mean for the routine use of EUS?

While the circumstances limiting the COGNATE trial to 223 randomisations were disappointing, only one of the 120 excluded because EUS was felt essential was in Aberdeen or Gloucester; these two centres therefore provide a representative picture of the care of gastro-oesophageal cancer under clinicians in equipoise. Thus COGNATE evaluated EUS wherever we could find equipoise about its value – specifically in two centres in equipoise and through individual clinicians in equipoise in six other centres.

Logically one would expect the other six centres to exclude from the COGNATE trial the patients considered likely to benefit from changed management plans following EUS; thus it would have been unethical for them to randomise these patients. In contrast, these centres were likely to randomise only those for whom there was uncertainty about likely benefit. This implies that the expected benefit of EUS for trial participants from those centres was less than for those whom they excluded from the COGNATE

trial. Nevertheless we found no significant differences between this virtual third centre and Aberdeen or Gloucester. If we respect the clinical judgement of those who made these decisions to exclude potential participants, we can infer that the expected benefit in those excluded was even greater than estimated by COGNATE. In short, this argument suggests that the benefit of EUS for gastro-oesophageal cancer may be even larger than the significant positive differences – in survival, quality-adjusted survival and cost-effectiveness – estimated by the COGNATE trial.

Referees interested in exploring these findings asked for subgroup analyses. Therefore we repeated survival analysis for several clinical subgroups and economic analysis for subgroups based on initial health status and age (see *Figures 25* and *26*); with the exception of sicker participants in *Figure 25*, we found effects to be less significant than in the full sample. Otherwise we avoided subgroup analyses because they reduce the power of comparisons: larger differences or smaller hazard ratios are needed to establish significance in a subgroup than in the full sample. Instead, if effects do differ significantly between subgroups, this will appear as an interaction with the relevant covariate in the full sample. The only such significant interaction was with baseline EQ-5D. There is no hint in any of our various multivariate analyses that centre, tumour site or tumour type had any influence on outcome. Given that ANCOVA is more powerful than subgroup analysis, we judge that these non-significant findings provide further support for the generalisability of our findings.

In summary, faced by difficulty in recruiting to a trial of a health technology no longer in equipoise in most centres across the UK, we decided to seek a smaller sample than originally planned – capable of detecting standardised differences of 0.4 rather than 0.3. Then COGNATE achieved statistical significance in favour of EUS on all three of our main criteria – survival, quality-adjusted survival and cost-effectiveness. Hence, as the concept of statistical power is prospective, this strategy overcame the undoubted prior weakness of a small sample.

Quality assurance

We undertook quality assurance – successfully despite difficulties of recruitment. Unfortunately the absence of a true gold standard with which to compare the consensual process limited the value of this work. Not only was no pathology available for patients not resected, but chemoradiotherapy may have downstaged the tumours of those resected. Furthermore few in this field would advocate using EUS to reassess tumours after chemoradiotherapy because of the difficulty of distinguishing residual tumour from the fibro-obliterative process that occurs with successful chemoradiotherapy.[71,160] Hence quality assurance within COGNATE did not address the whole process of care, but concentrated on scans and their interpretation. As EUS was the intervention, it was important to check the quality of reporting.

The role of quality assurance in pragmatic trials like COGNATE is to characterise 'best practice', ensure that the intervention achieves and maintains it, and thus reassure observers that the trial is robust. So it needs, not to influence the research, but to optimise best clinical practice. Hence it needs to be independent of the research but close enough to the clinical activity to perform this role. So we chose our panel of six from the centres who undertook the COGNATE trial and added an expert endosonographer from another continent. Although we had hoped to receive 40 usable videos and sample a random half, we reviewed the only 18 usable videos and added two sets of still images, one of which was from a centre without video equipment. Thus we achieved our aim of reviewing 20 (21%) of 97 scans performed in the EUS group. This needed five web-based and audio-facilitated consensus conferences of our panel over the course of the trial, remotely validated by the international assessor.

If the resulting opportunity sample of 20 scans was close to random, then we can be 95% confident that the overall proportion of acceptable scans lay between 87% and 100%. As the opportunity cost of these high-technology consensus conferences exceeded £10,000, we believe that the lower confidence bound

of 87% is sufficient for a pragmatic trial in which EUS also achieved significant benefits, not only in survival and quality-adjusted survival, but also in cost-effectiveness. Both the results and the consensus among contributors supported the proposition that quality of the scanning process within the COGNATE trial was reliable and generalisable.

Thus our prospective quality assurance programme achieved substantial consensus among six participating centres in inferring a common standard for reporting the TNM stages of gastro-oesophageal cancers through EUS. It reassured us about the quality of our intervention and helped us to meet the standards set for trials through GCP. It also fulfilled a useful educational role; for example, on finding they differed slightly in classifying borderline nodes, the panel agreed criteria for nodal involvement.

In short, we believe we improved consistency across centres and developed a model suitable for wider use after COGNATE. It is reassuring that other recent national trials, such as the US lung cancer screening trial reported by Gierada et al.,[161] have also monitored scans centrally. We therefore advocate central quality assurance for technologies like EUS, both in normal clinical practice and a fortiori in research.

Strengths and weaknesses relative to other studies

The only similar assessment of the clinical effects of EUS compared EUS FNA with a retrospective control group who did not receive EUS.[106] However that study neither randomised nor assessed the quality of EUS.

The value of quality assurance in clinical trials has been described in relatively few articles in the field of radiology. Toita et al.[162] designed a quality assurance of the process of delivering radiotherapy to ensure that the standards of treatment in their study complied with specific guidelines. Velasquez et al.[163] used a defined standard to ensure that PET scans were of adequate quality in the assessment of the value of PET scanning in the staging of gastrointestinal malignancies. Unfortunately we had no gold standard with which to compare our quality measures. Hence we used interobserver comparisons to measure the reliability of the staging investigations under study. In Germany Kutup et al.[98] used a 'gold standard' from tumours resected immediately after EUS: they correctly identified only two-thirds of T and N stages. However this restricted the applicability of their results to a small and unrepresentative proportion of gastro-oesophageal cancer. Similarly, fewer than 20% of participants assigned to EUS in the COGNATE trial had excision staging without intermediate chemotherapy.

The threshold of T4 as a reason for avoiding surgery is not as clear as some believe. Fockens et al.[113,164] have reported that in oesophageal carcinoma pleural involvement adjacent to the tumour is still physically resectable by removing the tumour en bloc with the overlying pleura; however the outcome for patients after surgical resection was no different from those not treated surgically. This was also evident in an American study by Chak et al.[165] However not all surgeons accept the lack of benefit for surgery in T4 as long as there is a learning curve in the interpretation of EUS in T4 disease in the oesophagus. For example Fockens et al.[113] reported that EUS underestimated T stage in 28 of a new EUS unit's first 100 examinations, and overestimated T stage in 14.

Endoscopic mucosal resection was planned in six COGNATE participants and undertaken in 12. The distinction between T1 and T2 is important only if this option is available. EMR is now an important clinical technique, especially for elderly patients with resectable T0 or T1 tumours but poor operative fitness. Centres in Europe and the USA have reported an increasing number.[166,167] The more advanced disease observed in this study is typical of the UK, which diagnoses little T0 or T1 disease and hence uses few conservative curative approaches like EMR and ablation.

Interpretation

Effectiveness

A combination of technical expertise and hard work enabled us to address many of the issues posed by a trial that was late and therefore small (see *Strengths and weaknesses of COGNATE*, *General*). Nevertheless COGNATE is not as generalisable as it would have been 10 years earlier. The effective study population comprised 70 participants from a teaching hospital in Aberdeen, 108 from a district general hospital in Gloucester, and a further 43 from centres that were not able to recruit in full. So it was reassuring to find no hint that the evidence from these three groups was at all heterogeneous. In particular we consistently tested: first whether the performance of EUS differed between the other centres and Aberdeen or Gloucester; and then whether it differed between Aberdeen and Gloucester; all without hint of statistical significance. Thus we are confident that COGNATE has estimated the true value of EUS in mainstream clinical practice. How could a trial that only just achieved minimal size achieve such clear benefits?

After EUS, one-quarter of management plans changed; participating clinicians explicitly attributed many changes in the intervention group to EUS. Although we had not expected much change in the control group, the proportion of changes in management was the same as in the intervention group. Those attributed to other diagnostic tests in the control group balanced those attributed to EUS in the intervention group, even though the results of most tests other than EUS had been available to the original MDM that had set the management plan. Given the encouraging findings about differences in outcome, notably survival, the apparent similarity in process was initially surprising.

Further analysis showed that participants allocated to EUS who then transferred to a therapeutic treatment, namely EMR or surgery in some form, survived much better than control participants who made this change, and better than intervention participants confirmed for one of these therapeutic treatments. The few intervention participants who transferred to conservative treatment – namely chemotherapy, radiotherapy or both – survived worse than control participants who made this change and worse than intervention participants confirmed for conservative treatment (see *Table 29*). In contrast, control participants who changed in either direction experienced intermediate survival, arguably because they lacked the discriminatory power of EUS. Furthermore intervention participants reported higher mean EQ-5D scores than the corresponding control participants. Finally, intervention participants who changed to multimodal treatments incurred lower costs than multimodal non-changers in the intervention group, possibly because many of them were already in advanced disease stages and hence died quickly; and intervention participants who changed to other treatments also incurred lower costs, arguably because their surgery or EMR was successful, leading to quicker recovery, less time in hospital and fewer recurrences. In short, changes were consistently more appropriate in the intervention group than in the control group.

Although all these unplanned analyses have low power, and are therefore not statistically significant, they stemmed from the need to explain significant differences in outcomes, yielded plausible reasons for those differences, and helped to explain them. They also yielded plausible explanations for the similar proportions of management changes in each group: changes arising from EUS were larger in effect and more discriminatory in nature. The better resection rates in the intervention group, even among those whose plans remain unaltered, confirm these observations. Our quality assurance data are consistent with the propositions that EUS triggers switches to better treatment and also helps surgeons undertaking resection and oncologists devising multimodal treatment without changing their basic plans. Furthermore our data on the use of NHS resources in both groups show that, although the findings of EUS usually led to action without delay, its absence led to additional tests and extended average stay by one-third. Although we have no data on the confidence of clinicians and participants in both groups, it is possible that some of the benefits of EUS stemmed from the psychological boost of prompt discharge, in addition to technical improvements in staging, management plans and resection.

The higher proportion of resections completed in the intervention group suggests that EUS improves selection for surgery by finding more resectable tumours. If so, that is one biological reason why EUS has a beneficial effect on survival and quality of life. The way in which changes are consistently more appropriate in the intervention group than in the control group suggests a more pervasive effect: even when EUS does not lead to a change of plan, it nevertheless enables clinicians to confirm their chosen plan, and helps them to implement it; and they convey the resulting increase in confidence to their patients. As this was a pragmatic trial, however, it did not directly address the mechanisms by which EUS might improve survival and quality of life. Our supplementary indirect analyses help to fill this gap.

Cost-effectiveness

Since 2008, NICE has recommended that the appraisal of health technology should consider benefits only to patients and resource costs only to the NHS.[148] Many have argued that benefits and costs to others, notably patients' families, are equally important in such appraisals. In recent elegant economic analysis, however, Claxton *et al.*[168] have underpinned the NICE policy, showing that for them to ignore the natural constraint on the NHS budget would lead to greater distortion in public decision-making than it corrects. Although we are conscious that in the field of cancer the NICE policy does generate anomalies, we have been happy to follow their prudent advice, not least because it simplifies analysis.

To cost NHS resources we used published national unit costs and refined these in the light of detailed consultation with the Nuclear Medicine and Finance Departments of Grampian Universities NHS Trust. This led us to estimate the cost of EUS as £550 for day patients, £1500 for outpatients and £3800 for inpatients. Hence we used sensitivity analysis to test whether this wide variation affected our conclusion about the cost-effectiveness of EUS.

The National Institute of Health and Care Excellence has suggested that interventions costing less than a threshold of between £20,000 and £30,000 (depending on circumstances) per QALY are likely to use NHS resources cost-effectively.[148] At all thresholds in this range EUS has a more than 95% chance of being cost-effective. Given that EUS appears to save costs, we also considered a threshold of £0 per QALY, which implies that the NHS is unwilling to pay for extra QALYs but insists that innovations such as EUS should finance themselves: even then it has more than 90% chance of being cost-effective. Finally we considered an infinite threshold per QALY – even that suggests that EUS is likely to be cost-effective (85% chance).

We also found significant interaction in cost-effectiveness analyses: participants with low baseline EQ-5D scores gained an average of 0.31 QALYs from EUS, whereas those with high baseline EQ-5D scores gained an average of only 0.018 QALYs (see *Table 42*). Nevertheless detailed analysis shows that EUS is probably cost-effective even for these healthier participants because they have lower costs. So there is no scientific reason to restrict access to EUS to patients who are less fit.

Synthesis

We judge that the consistency of our findings provides convincing evidence that EUS is effective and cost-effective. As we achieved the minimal sample size for our primary outcome, but only some secondary outcomes, we could not achieve statistical significance throughout. Nevertheless we judge that the combination of: increasing the complete resection rate from 80% to 91%; significant improvements in survival and quality-adjusted survival; and consistent improvements in patient-reported outcomes, in survival across subgroups defined by changes in management plans, and in resource use – all resulting in 95% probability that EUS is cost-effective – is conclusive.

Future research

Metaphorically, the COGNATE team caught the EUS horse just as it was leaving the stable. In these circumstances we doubt whether rigorous research to refine the findings of the COGNATE trial is feasible. So we make no recommendation for future research into EUS. Instead, given the considerable difficulty we experienced in completing the COGNATE trial successfully, the main contribution of this section is to emphasise the need for research into the best time to evaluate new technologies. Given the recent success of a range of schemes to facilitate the evaluation of expensive pharmaceuticals, including 'coverage for evidence development', 'only in research', 'patient access schemes', and 'risk sharing' between manufacturer and NHS (e.g. Hughes *et al*.[169]), we suggest that similar schemes may have merit in expediting the evaluation of expensive health technologies (in the widest sense), especially if they claim to ameliorate relatively rare conditions. Whatever the method of funding such technologies and their evaluation, however, there is a strong case for effective 'horizon scanning' to identify them while evaluation is still feasible.

The COGNATE trial gives rise to two strong recommendations for methodological research. First, we advocate further work to refine electronic and other methods of streamlining the collection of data to evaluate complex technologies like EUS, notably on the costs of the extensive care for conditions such as gastro-oesophageal cancer. Secondly, we are keen to refine the integrated FACT Oesophageal and Gastric scale (FACT-EG) as a valid measure of the outcome of gastro-oesophageal cancer.

Chapter 5 Conclusions

It is very difficult to assess health technologies that have already been widely adopted. We achieved a rigorous evaluation of EUS in staging gastro-oesophageal cancer only because:

- We twice reduced our target sample size, first by combining survival and quality of life to form a composite primary outcome in the form of quality-adjusted survival, and secondly by increasing our target effect size from 0.3 to 0.4.
- The NIHR HTA programme kindly granted no-cost extensions of 6 months to complete recruitment, and 12 months to complete this report in the face of difficult personal circumstances.
- Two centres – Aberdeen and Gloucester – remained in equipoise about the practical value of EUS, and recruited participants enthusiastically and effectively over 3.5 years.
- Another six centres allowed individual clinicians in equipoise to recruit participants as and when they could.
- The research team worked hard to maintain scientific rigour in the face of adversity.

Against this unpromising background EUS achieved a surprising combination of positive findings in gastro-oesophageal cancer patients from centres in equipoise and those for whom clinicians from other centres feel that EUS is not mandatory:

- Significant improvement in survival – summarised by a hazard ratio of 0.706 (95% CI from 0.501 to 0.996) and an increase of 121 days in estimated median survival (from 1.63 years in the control group to 1.96 years in the EUS group).
- Consistent though individually non-significant improvements in mean participant-reported outcomes at 12 months – characterised by a difference of 0.060 (95% CI from –0.041 to 0.161) in mean EQ-5D scores between 0.444 in the control group and 0.503 in the intervention group, and by a difference of 0.12 (95% CI from –0.27 to 0.51) in mean FACT-G cancer-specific scores between 2.15 in the control group and 2.27 in the EUS group.
- Significant improvement in quality-adjusted survival (i.e. adjusted for generic quality of life, namely EQ-5D) – summarised by a hazard ratio of 0.705 (95% from CI 0.499 to 0.995) and an increase of 66 days in median quality-adjusted survival – from 0.94 QALYs in the control group to 1.12 QALYs in the intervention group.
- Significant contrast between initially sicker and healthier participants, with EUS providing most benefit to those who are sicker.
- Consistent, though non-significant, reductions in total resource use over 48 months in secondary and pharmaceutical care (including the estimated cost of EUS scans) – summarised by a mean savings of about £2860 (95% 'bootstrapped' CI from –£2200 to £8000) from an average of £32,000 (SD £22,000) in the control group to £29,200 (SD £14,900) in the intervention group.
- Probability of 96.6% that EUS is cost-effective in the sense of achieving the stricter NICE threshold of costing less than a mean of £20,000 to gain a QALY.
- Increase in the complete resection rate from 80% (44 successes out of 55 attempts in the control group) to 91% (48 successes out of 53 attempts in the intervention group).
- Changes in management plans that were consistently more appropriate in the group allocated to EUS than in the control group.

In summary, COGNATE was some 10 years too late, because EUS was common practice by the time it began recruiting in 2005. Nevertheless the COGNATE team ameliorated many of the resulting problems of recruiting centres and participants, mainly through the efforts of two centres – Aberdeen and Gloucester. Moreover EUS achieved an impressive combination of positive results; those for all primary outcomes were statistically significant. We judge that these findings provide strong evidence in favour of EUS scans for all gastro-oesophageal cancer patients with the potential to benefit.

Acknowledgements

We take great pleasure in thanking all whose commitment and effort enabled the COGNATE team to submit this final report: participants at the eight contributing hospitals for providing data in difficult circumstances; PIs, clinical and research staff at the hospitals named in *Appendix 2,* especially Maureen Gillan, Heather Hodgson and Clive Stokes, for their assiduous work for a trial seeking rigorous evidence in the face of criticism that it was unethical and unscientific; our quality assurance team, notably Dympna McAteer and Peter Rodgers, for diligent contributions, and Professor KM Mahondas for loyally acting as external reviewer; the trial team listed in *Appendix 3* and the other members of the Institute of Medical and Social Care Research at Bangor University, notably Dyfrig Hughes, Karen Hughes, Kevin Mawdesley, Huw Roberts and Chris Whitaker, for generous technical, administrative and affective support; members of the TSC, especially Marion Campbell, Hugh Gilmour, Robert Heading, Tony Lerut and David Kirby for their principled and reassuring guidance in stressful times; the NCCHTA for generous funding; and their congenial and effective staff for consistent and sympathetic administration of a trial that encountered more adversity than most.

Contribution of authors

Ian Russell, trialist, was methodological lead of COGNATE and edited the final draft of this report.

Professor Rhiannon Tudor Edwards, health economist, designed and led the health economic component of COGNATE.

Dr Angela Gliddon was trial manager of COGNATE and co-ordinated the first draft of this report.

Dr David Ingledew, senior lecturer in psychology, designed and led the quality-of-life component of COGNATE.

Dr Daphne Russell, senior trial statistician, led the statistical component of COGNATE, analysed the effectiveness data, and edited the final draft of this report.

Ms Rhiannon Whitaker, statistician and assistant director of the Bangor Trials Unit, managed and analysed the quality-of-life data, and edited the first draft of this report.

Ms Seow Tien Yeo, research officer in health economics, managed and analysed the health economic data.

Mr Stephen Attwood, consultant surgeon, led the quality assurance component of COGNATE.

Professor Hugh Barr, consultant surgeon, was PI in the Gloucester centre, which recruited more than 100 patients.

Mr Shayanthan Nanthakumaran, specialist registrar in surgery, undertook the literature review reported in *Chapter 1*.

Professor Ken Park, consultant surgeon, was clinical lead of COGNATE and PI in the Aberdeen centre, which recruited more than 70 patients.

References

1. Cancer Research UK. *Oesophageal Cancer Incidence Statistics.* http://info.cancerresearchuk.org/cancerstats/types/oesophagus/incidence (accessed August 2011).

2. Cancer Research UK. *Stomach cancer incidence statistics.* http://info.cancerresearchuk.org/cancerstats/types/stomach/incidence (accessed August 2011).

3. Dolan K, Sutton R, Walker SJ, Morris AI, Campbell F, Williams EM. New classification of oesophageal and gastric carcinomas derived from changing patterns in epidemiology. *Br J Cancer* 1999;**80**:834–42. http://dx.doi.org/10.1038/sj.bjc.6690429

4. Gilbert FJ, Park KGM, Thompson AM, Rapson T, Thomson CS. *Scottish audit of gastric and oesophageal cancer: report 1997–2000.* Edinburgh: Scottish Executive Health Department; 2002.

5. Palser T, Cromwell D, van der Meulen J, Hardwick R, Riley S, Greenaway K, *et al. National oesophago-gastric cancer audit: first annual report.* Leeds: NHS Information Centre; 2008.

6. Office for National Statistics. *Mortality statistics: review by Registrar General of deaths by cause*, *sex and age in England and Wales in 2005.* London: Her Majesty's Stationery Office; 2006.

7. Ahmadi A, Draganov P. Endoscopic mucosal resection in the upper gastrointestinal tract. *World J Gastroenterol* 2008;**14**:1984–9. http://dx.doi.org/10.3748/wjg.14.1984

8. Uedo N, Iishi H, Tatsuta M, Ishihara R, Higashino K, Takeuchi Y, *et al.* Longterm outcomes after endoscopic mucosal resection for early gastric cancer. *Gastric Cancer* 2006;**9**:88–92. http://dx.doi.org/10.1007/s10120-005-0357-0

9. Araki K, Ohno S, Egashira A, Saeki H, Kawaguchi H, Sugimachi K. Pathologic features of superficial esophageal squamous cell carcinoma with lymph node and distal metastasis. *Cancer* 2002;**94**:570–5. http://dx.doi.org/10.1002/cncr.10190

10. Gotoda T. Endoscopic resection of early gastric cancer. *Gastric Cancer* 2007;**10**:1–11. http://dx.doi.org/10.1007/s10120-006-0408-1

11. Higuchi K, Tanabe S, Koizumi W, Sasaki T, Nakatani K, Saigenji K, *et al.* Expansion of the indications for endoscopic mucosal resection in patients with superficial esophageal carcinoma. *Endoscopy* 2007;**39**:36–40. http://dx.doi.org/10.1055/s-2006-945148

12. Blazeby JM, Metcalfe C, Nicklin J, Barham CP, Donovan J, Alderson D. Association between quality of life scores and short-term outcome after surgery for cancer of the oesophagus or gastric cardia. *Br J Surg* 2005;**92**:1502–7. http://dx.doi.org/10.1002/bjs.5175

13. Al-Sarira AA, David G, Willmott S, Slavin JP, Deakin M, Corless DJ. Oesophagectomy practice and outcomes in England. *Br J Surg* 2007;**94**:585–91. http://dx.doi.org/10.1002/bjs.5805

14. De Boer AG, Genovesi PI, Sprangers MA, Van Sandick JW, Obertop H, Van Lanschot JJ. Quality of life in long-term survivors after curative transhiatal oesophagectomy for oesophageal carcinoma. *Br J Surg* 2000;**87**:1716–21. http://dx.doi.org/10.1046/j.1365-2168.2000.01600.x

15. Dexter SP, Sue-Ling H, McMahon MJ, Quirke P, Mapstone N, Martin IG. Circumferential resection margin involvement: an independent predictor of survival following surgery for oesophageal cancer. *Gut* 2001;**48**:667–70. http://dx.doi.org/10.1136/gut.48.5.667

16. Suttie SA, Li AG, Quinn M, Park KG. The impact of operative approach on outcome of surgery for gastro-oesophageal tumours. *World J Surg Oncol* 2007;**5**:95. http://dx.doi.org/10.1186/1477-7819-5-95

17. Vogt K, Fenlon D, Rhodes S, Malthaner RA. Preoperative chemotherapy for resectable thoracic esophageal cancer. *Cochrane Database Syst Rev* 2006;**3**:CD001556.

18. Medical Research Council Oesophageal Cancer Working Group. Surgical resection with or without preoperative chemotherapy in oesophageal cancer: a randomised controlled trial. *Lancet* 2002;**359**:1727–33. http://dx.doi.org/10.1016/S0140-6736(02)08651-8

19. Kelsen DP, Ginsberg R, Pajak TF, Sheahan DG, Gunderson L, Mortimer J, *et al.* Chemotherapy followed by surgery compared with surgery alone for localized esophageal cancer. *N Engl J Med* 1998;**339**:1979–84. http://dx.doi.org/10.1056/NEJM199812313392704

20. Paoletti X, Oba K, Burzykowski T, Michiels S, Ohashi Y, Pignon JP, *et al.* Benefit of adjuvant chemotherapy for resectable gastric cancer: a meta-analysis. *JAMA* 2010;**303**:1729–37. http://dx.doi.org/10.1001/jama.2010.534

21. Macdonald JS, Smalley SR, Benedetti J, Hundahl SA, Estes NC, Stemmermann GN, *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. http://dx.doi.org/10.1056/NEJMoa010187

22. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJH, Nicolson M, *et al.* Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med* 2006;**355**:11–20. http://dx.doi.org/10.1056/NEJMoa055531

23. Earlam R, Cunha-Melo JR. Oesophogeal squamous cell carcinomas: II. critical view of radiotherapy. *Br J Surg* 1980;**67**:457–61. http://dx.doi.org/10.1002/bjs.1800670702

24. Sykes AJ, Burt PA, Slevin NJ, Stout R, Marrs JE. Radical radiotherapy for carcinoma of the oesophagus: an effective alternative to surgery. *Radiother Oncol* 1998;**48**:15–21. http://dx.doi.org/10.1016/S0167-8140(98)00037-1

25. Okawa T, Kita M, Tanaka M, Ikeda M. Results of radiotherapy for inoperable locally advanced esophageal cancer. *Int J Radiat Oncol Biol Phys* 1989;**17**:49–54. http://dx.doi.org/10.1016/0360-3016(89)90369-6

26. Crosby TD, Brewster AE, Borley A, Perschky L, Kehagioglou P, Court J, *et al.* Definitive chemoradiation in patients with inoperable oesophageal carcinoma. *Br J Cancer* 2004;**90**:70–5. http://dx.doi.org/10.1038/sj.bjc.6601461

27. Morgan MA, Lewis WG, Casbard A, Roberts SA, Adams R, Clark GW, *et al.* Stage-for-stage comparison of definitive chemoradiotherapy, surgery alone and neoadjuvant chemotherapy for oesophageal carcinoma. *Br J Surg* 2009;**96**:1300–7. http://dx.doi.org/10.1002/bjs.6705

28. Cooper JS, Guo MD, Herskovic A, Macdonald JS, Martenson JA, Al-Sarraf M, *et al.* Chemoradiotherapy of locally advanced esophageal cancer: long-term follow-up of a prospective randomized trial. *JAMA* 1999;**281**:1623–7. http://dx.doi.org/10.1001/jama.281.17.1623

29. Bedenne L, Michel P, Bouche O, Milan C, Mariette C, Conroy T, *et al.* Chemoradiation followed by surgery compared with chemoradiation alone in squamous cancer of the esophagus. *J Clin Oncol* 2007;**25**:1160–8. http://dx.doi.org/10.1200/JCO.2005.04.7118

30. Stahl M, Stuschke M, Lehmann N, Meyer HJ, Walz MK, Seeber S, *et al.* Chemoradiation with and without surgery in patients with locally advanced squamous cell carcinoma of the esophagus. *J Clin Oncol* 2005;**23**:2310–17. http://dx.doi.org/10.1200/JCO.2005.00.034

31. Chiu PW, Chan AC, Leung SF, Leong HT, Kwong KH, Li MK, *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. http://dx.doi.org/10.1016/j.gassur.2005.05.005

32. Herskovic A, Martz K, al-Sarraf M, Leichman L, Brindle J, Vaitkevicius V, *et al.* Combined chemotherapy and radiotherapy compared with radiotherapy alone in patients with cancer of the esophagus. *N Engl J Med* 1992;**326**:1593–8. http://dx.doi.org/10.1056/NEJM199206113262403

33. Wong R, Malthaner R. Combined chemotherapy and radiotherapy (without surgery) compared with radiotherapy alone in localized carcinoma of the esophagus2010. *Cochrane Database Syst Rev* 2010;**3**:CD002092.

34. Kelly S, Harris KM, Berry E, Hutton J, Roderick P, Cullingworth J, *et al.* A systematic review of the staging performance of endoscopic ultrasound in gastro-oesophageal carcinoma. *Gut* 2001;**49**:534–9. http://dx.doi.org/10.1136/gut.49.4.534

35. 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**:1479–90. http://dx.doi.org/10.3748/wjg.14.1479

36. Puli SR, Reddy JBK, Bechtold ML, Antillon MR, Ibdah JA. How good is endoscopic ultrasound for TNM staging of gastric cancers? A meta-analysis and systematic review. *World J Gastroenterol* 2008;**14**:4011–19. http://dx.doi.org/10.3748/wjg.14.4011

37. Ahn HS, Lee HJ, Yoo MW, Kim SG, Im JP, Kim SH, *et al.* Diagnostic accuracy of T and N stages with endoscopy, stomach protocol CT, and endoscopic ultrasonography in early gastric cancer. *J Surg Oncol* 2009;**99**:20–7. http://dx.doi.org/10.1002/jso.21170

38. Akahoshi K, Misawa T, Fujishima H, Chijiiwa Y, Nawata H. Regional lymph node metastasis in gastric cancer: evaluation with endoscopic US. *Radiology* 1992;**182**:559–64.

39. Ang TL, Ng TM, Fock KM, Teo EK. Accuracy of endoscopic ultrasound staging of gastric cancer in routine clinical practice in Singapore. *Chin J Dig Dis* 2006;**7**:191–6. http://dx.doi.org/10.1111/j.1443-9573.2006.00270.x

40. Arocena MG, Barturen A, Bujanda L, Casado O, Ramirez MM, Oleagoitia JM, *et al.* MRI and endoscopic ultrasonography in the staging of gastric cancer. *Rev Esp Enferm Dig* 2006;**98**:582–90. http://dx.doi.org/10.4321/S1130-01082006000800003

41. Bentrem D, Gerdes H, Tang L, Brennan M, Coit D. Clinical correlation of endoscopic ultrasonography with pathologic stage and outcome in patients undergoing curative resection for gastric cancer. *Ann Surg Oncol* 2007;**14**:1853–9. http://dx.doi.org/10.1245/s10434-006-9037-5

42. Bhandari S, Shim CS, Kim JH, Jung IS, Cho JY, Lee JS, *et al.* Usefulness of three-dimensional, multidetector row CT (virtual gastroscopy and multiplanar reconstruction) in the evaluation of gastric cancer: a comparison with conventional endoscopy, EUS and histopathology. *Gastrointest Endosc* 2004;**59**:619–26. http://dx.doi.org/10.1016/S0016-5107(04)00169-5

43. Botet JF, Lightdale CJ, Zauber AG, Gerdes H, Winawer SJ, Urmacher C, *et al.* Preoperative staging of gastric cancer: comparison of endoscopic US and dynamic CT. *Radiology* 1991;**181**:426–32.

44. Caletti G, Ferrari A, Brocchi E, Barbara L. Accuracy of endoscopic ultrasonography in the diagnosis and staging of gastric cancer and lymphoma. *Surgery* 1993;**113**:14–27.

45. Dittler HJ, Siewert JR. Role of endoscopic ultrasonography in gastric carcinoma. *Endoscopy* 1993;**25**:162–6. http://dx.doi.org/10.1055/s-2007-1010276

46. Francois E, Peroux J, Mouroux J, Chazalle M, Hastier P, Ferrero J, *et al.* Preoperative endosonographic staging of cancer of the cardia. *Abdom Imaging* 1996;**21**:483–7. http://dx.doi.org/10.1007/s002619900109

47. Ganpathi IS, So JB, Ho KY. Endoscopic ultrasonography for gastric cancer: does it influence treatment? *Surg Endosc* 2006;**20**:559–62. http://dx.doi.org/10.1007/s00464-005-0309-0

48. Grimm H, Binmoeller KF, Hamper K, Koch J, Henne-Bruns D, Soehendra N. Endosonography for preoperative locoregional staging of esophageal and gastric cancer. *Endoscopy* 1993;**25**:224–30. http://dx.doi.org/10.1055/s-2007-1010297

49. Hunerbein M, Dohmoto M, Rau B, Schlag PM. Endosonography and endosonography-guided biopsy of upper-GI-tract tumors using a curved-array echoendoscope. *Surg Endosc* 1996;**10**:1205–9. http://dx.doi.org/10.1007/s004649900280

50. Javaid G, Shah OJ, Dar MA, Shah P, Wani NA, Zargar SA. Role of endoscopic ultrasonography in preoperative staging of gastric carcinoma. *Aust NZ J Surg* 2004;**74**:108–11. http://dx.doi.org/10.1046/j.1445-1433.2003.02923.x

51. Lok KH, Lee CK, Yiu HL, Lai L, Szeto ML, Leung SK. Current utilization and performance status of endoscopic ultrasound in a community hospital. *Chin J Dig Dis* 2008;**9**:41–7. http://dx.doi.org/10.1111/j.1443-9573.2007.00318.x

52. Massari M, Cioffi U, De Simone M, Bonavina L, D'Elia A, Rosso L, *et al.* Endoscopic ultrasonography for preoperative staging of gastric carcinoma. *Hepatogastroenterology* 1996;**43**:542–6.

53. Murata Y, Suzuki S, Hashimoto H. Endoscopic ultrasonography of the upper gastrointestinal tract. *Surg Endosc* 1988;**2**:180–3. http://dx.doi.org/10.1007/BF02498796

54. Perng DS, Jan CM, Wang WM, Chen LT, Su YC, Liu GC, *et al.* Computed tomography, endoscopic ultrasonography and intraoperative assessment in TN staging of gastric carcinoma. *J Formos Med Assoc* 1996;**95**:378–85.

55. Polkowski M, Palucki J, Wronska E, Szawlowski A, Nasierowska-Guttmejer A, Butruk E. Endosonography versus helical computed tomography for locoregional staging of gastric cancer. *Endoscopy* 2004;**36**:617–23. http://dx.doi.org/10.1055/s-2004-814522

56. Potrc S, Skalicky M, Ivanecz A. Does endoscopic ultrasound staging already allow individual treatment regimens in gastric cancer. *Wien Klin Wochenschr* 2006;**118**(Suppl. 2):48–51. http://dx.doi.org/10.1007/s00508-006-0552-y

57. Saito N, Takeshita K, Habu H, Endo M. The use of endoscopic ultrasound in determining the depth of cancer invasion in patients with gastric cancer. *Surg Endosc* 1991;**5**:14–19. http://dx.doi.org/10.1007/BF00591380

58. Shimoyama S, Yasuda H, Hashimoto M, Tatsutomi Y, Aoki F, Mafune K, *et al.* Accuracy of linear-array EUS for preoperative staging of gastric cardia cancer. *Gastrointest Endosc* 2004;**60**:50–5. http://dx.doi.org/10.1016/S0016-5107(04)01312-4

59. Tio TL, Schouwink MH, Cikot RJ, Tytgat GN. Preoperative TNM classification of gastric carcinoma by endosonography in comparison with the pathological TNM system: a prospective study of 72 cases. *Hepatogastroenterology* 1989;**36**:51–6.

60. Tsendsuren T, Jun SM, Mian XH. Usefulness of endoscopic ultrasonography in preoperative TNM staging of gastric cancer. *World J Gastroenterol* 2006;**12**:43–7.

61. Wang JY, Hsieh JS, Huang YS, Huang CJ, Hou MF, Huang TJ. Endoscopic ultrasonography for preoperative locoregional staging and assessment of resectability in gastric cancer. *Clin Imaging* 1998;**22**:355–9. http://dx.doi.org/10.1016/S0899-7071(98)00033-3

62. Willis S, Truong S, Gribnitz S, Fass J, Schumpelick V. Endoscopic ultrasonography in the preoperative staging of gastric cancer: accuracy and impact on surgical therapy. *Surg Endosc* 2000;**14**:951–4. http://dx.doi.org/10.1007/s004640010040

63. Xi WD, Zhao C, Ren GS. Endoscopic ultrasonography in preoperative staging of gastric cancer: determination of tumor invasion depth, nodal involvement and surgical resectability. *World J Gastroenterol* 2003;**9**:254–7.

64. Ziegler K, Sanft C, Zimmer T, Zeitz M, Felsenberg D, Stein H, *et al.* Comparison of computed tomography, endosonography and intraoperative assessment in TN staging of gastric carcinoma. *Gut* 1993;**34**:604–10. http://dx.doi.org/10.1136/gut.34.5.604

65. Rosch T, Lorenz R, Zenker K, von Wichert A, Dancygier H, Hofler H, *et al.* Local staging and assessment of resectability in carcinoma of the esophagus, stomach, and duodenum by endoscopic ultrasonography. *Gastrointest Endosc* 1992;**38**:460–7. http://dx.doi.org/10.1016/S0016-5107(92)70477-5

66. Kwee RM, Kwee TC. Imaging in assessing lymph node status in gastric cancer. *Gastric Cancer* 2009;**12**:6–22. http://dx.doi.org/10.1007/s10120-008-0492-5

67. Kwee RM, Kwee TC. Imaging in local staging of gastric cancer: a systematic review. *J Clin Oncol* 2007;**25**:2107–16. http://dx.doi.org/10.1200/JCO.2006.09.5224

68. Kwee RM, Kwee TC. The accuracy of endoscopic ultrasonography in differentiating mucosal from deeper gastric cancer. *Am J Gastroenterol* 2008;**103**:1801–9. http://dx.doi.org/10.1111/j.1572-0241.2008.01923.x

69. Vilgrain V, Mompoint D, Palazzo L, Menu Y, Gayet B, Ollier P, *et al.* Staging of esophageal carcinoma: comparison of results with endoscopic sonography and CT. *Am J Roentgenol* 1990;**155**:277–81.

70. Botet JF, Lightdale CJ, Zauber AG, Gerdes H, Urmacher C, Brennan MF. Preoperative staging of esophageal cancer: comparison of endoscopic US and dynamic CT. *Radiology* 1991;**181**:419–25.

71. Rice TW, Boyce GA, Sivak MV. Esophageal ultrasound and the preoperative staging of carcinoma of the esophagus. *J Thorac Cardiovasc Surg* 1991;**101**:536–44. 71

72. Ziegler K, Sanft C, Zeitz M, Friedrich M, Stein H, Haring R, *et al.* Evaluation of endosonography in TN staging of oesophageal cancer. *Gut* 1991;**32**:16–20. http://dx.doi.org/10.1136/gut.32.1.16

73. Dittler HJ, Siewert JR. Role of endoscopic ultrasonography in esophageal carcinoma. *Endoscopy* 1993;**25**:156–61. http://dx.doi.org/10.1055/s-2007-1010275

74. Hordijk ML, Zander H, van Blankenstein M, Tilanus HW. Influence of tumor stenosis on the accuracy of endosonography in preoperative T staging of esophageal cancer. *Endoscopy* 1993;**25**:171–5. http://dx.doi.org/10.1055/s-2007-1010278

75. Greenberg J, Durkin M, Van Drunen M, Aranha GV. Computed tomography or endoscopic ultrasonography in preoperative staging of gastric and esophageal tumors. *Surgery* 1994;**116**:696–702.

76. Peters JH, Hoeft SF, Heimbucher J, Bremner RM, DeMeester TR, Bremner CG, *et al.* Selection of patients for curative or palliative resection of esophageal cancer based on preoperative endoscopic ultrasonography. *Arch Surg* 1994;**129**:534–9. http://dx.doi.org/10.1001/archsurg.1994.01420290080012

77. Yoshikane H, Tsukamoto Y, Niwa Y, Goto H, Hase S, Shimodaira M, *et al.* Superficial esophageal carcinoma: evaluation by endoscopic ultrasonography. *Am J Gastroenterol* 1994;**89**:702–7.

78. Binmoeller KF, Seifert H, Seitz U, Izbicki JR, Kida M, Soehendra N. Ultrasonic esophagoprobe for TNM staging of highly stenosing esophageal carcinoma. *Gastrointest Endosc* 1995;**41**:547–52. http://dx.doi.org/10.1016/S0016-5107(95)70188-5

79. McLoughlin RF, Cooperberg PL, Mathieson JR, Stordy SN, Halparin LS. High resolution endoluminal ultrasonography in the staging of esophageal carcinoma. *J Ultrasound Med* 1995;**14**:725–30.

80. Hasegawa N, Niwa Y, Arisawa T, Hase S, Goto H, Hayakawa T. Preoperative staging of superficial esophageal carcinoma: comparison of an ultrasound probe and standard

endoscopic ultrasonography. *Gastrointest Endosc* 1996;**44**:388–93. http://dx.doi.org/10.1016/S0016-5107(96)70086-X

81. Holden A, Mendelson R, Edmunds S. Pre-operative staging of gastro-oesophageal junction carcinoma: comparison of endoscopic ultrasound and computed tomography. *Australas Radiol* 1996;**40**:206–12. http://dx.doi.org/10.1111/j.1440-1673.1996.tb00386.x

82. Natsugoe S, Yoshinaka H, Morinaga T, Shimada M, Baba M, Fukumoto T, *et al.* Ultrasonographic detection of lymph-node metastases in superficial carcinoma of the esophagus. *Endoscopy* 1996;**28**:674–9. http://dx.doi.org/10.1055/s-2007-1005575

83. Massari M, Cioffi U, De Simone M, Lattuada E, Montorsi M, Segalin A, *et al.* Endoscopic ultrasonography for preoperative staging of esophageal carcinoma. *Surg Laparosc Endosc* 1997;**7**:162–5. http://dx.doi.org/10.1097/00019509-199704000-00021

84. Pham T, Roach E, Falk GL, Chu J, Ngu MC, Jones DB. Staging of oesophageal carcinoma by endoscopic ultrasound: preliminary experience. *Aust NZ J Surg* 1998;**68**:209–12. http://dx.doi.org/10.1111/j.1445-2197.1998.tb04748.x

85. Vickers J, Alderson D. Oesophageal cancer staging using endoscopic ultrasonography. *Br J Surg* 1998;**85**:994–8. http://dx.doi.org/10.1046/j.1365-2168.1998.00694.x

86. Bowrey DJ, Clark GW, Roberts SA, Maughan TS, Hawthorne AB, Williams GT, *et al.* Endosonographic staging of 100 consecutive patients with esophageal carcinoma: introduction of the 8-mm esophagoprobe. *Dis Esophagus* 1999;**12**:258–63. http://dx.doi.org/10.1046/j.1442-2050.1999.00071.x

87. Catalano MF, Alcocer E, Chak A, Nguyen CC, Raijman I, Geenen JE, *et al.* Evaluation of metastatic celiac axis lymph nodes in patients with esophageal carcinoma: accuracy of EUS. *Gastrointest Endosc* 1999;**50**:352–6. http://dx.doi.org/10.1053/ge.1999.v50.98154

88. Nishimaki T, Tanaka O, Ando N, Ide H, Watanabe H, Shinoda M, *et al.* Evaluation of the accuracy of preoperative staging in thoracic esophageal cancer. *Ann Thorac Surg* 1999;**68**:2059–64. http://dx.doi.org/10.1016/S0003-4975(99)01171-6

89. Salminen JT, Farkkila MA, Ramo OJ, Toikkanen V, Simpanen J, Nuutinen H, *et al.* Endoscopic ultrasonography in the preoperative staging of adenocarcinoma of the distal oesophagus and oesophagogastric junction. *Scand J Gastroenterol* 1999;**34**:1178–82. http://dx.doi.org/10.1080/003655299750024670

90. Heidemann J, Schilling MK, Schmassmann A, Maurer CA, Buchler MW. Accuracy of endoscopic ultrasonography in preoperative staging of esophageal carcinoma. *Dig Surg* 2000;**17**:219–24. http://dx.doi.org/10.1159/000018838

91. Nesje LB, Svanes K, Viste A, Laerum OD, Odegaard S. Comparison of a linear miniature ultrasound probe and a radial-scanning echoendoscope in TN staging of esophageal cancer. *Scand J Gastroenterol* 2000;**35**:997–1002. http://dx.doi.org/10.1080/003655200750023101

92. Vazquez-Sequeiros E, Norton ID, Clain JE, Wang KK, Affi A, Allen M, *et al.* Impact of EUS-guided fine-needle aspiration on lymph node staging in patients with esophageal carcinoma. *Gastrointest Endosc* 2001;**53**:751–7. http://dx.doi.org/10.1067/mge.2001.112741

93. Kienle P, Buhl K, Kuntz C, Dux M, Hartmann C, Axel B, *et al.* Prospective comparison of endoscopy, endosonography and computed tomography for staging of tumours of the oesophagus and gastric cardia. *Digestion* 2002;**66**:230–6. http://dx.doi.org/10.1159/000068360

94. Chang KJ, Soetikno RM, Bastas D, Tu C, Nguyen PT. Impact of endoscopic ultrasound combined with fine-needle aspiration biopsy in the management of esophageal cancer. *Endoscopy* 2003;**35**:962–6. http://dx.doi.org/10.1055/s-2003-43470

95. Wu LF, Wang BZ, Feng JL, Cheng WR, Liu GR, Xu XH, *et al.* Preoperative TN staging of esophageal cancer: comparison of miniprobe ultrasonography, spiral CT and MRI. *World J Gastroenterol* 2003;**9**:219–24.

96. DeWitt J, Kesler K, Brooks JA, LeBlanc J, McHenry L, McGreevy K, *et al.* Endoscopic ultrasound for esophageal and gastroesophageal junction cancer: impact of increased use of primary neoadjuvant therapy on preoperative locoregional staging accuracy. *Dis Esophagus* 2005;**18**:21–7. http://dx.doi.org/10.1111/j.1442-2050.2005.00444.x

97. 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**:422–30. http://dx.doi.org/10.1007/s11307-005-0017-0

98. Kutup A, Link BC, Schurr PG, Strate T, Kaifi JT, Bubenheim M, *et al.* Quality control of endoscopic ultrasound in preoperative staging of esophageal cancer. *Endoscopy* 2007;**39**:715–19. http://dx.doi.org/10.1055/s-2007-966655

99. Moorjani N, Junemann-Ramirez M, Judd O, Fox B, Rahamim JS. Endoscopic ultrasound in esophageal carcinoma: comparison with multislice computed tomography and importance in the clinical decision making process. *Minerva Chir* 2007;**62**:217–23.

100. Shimpi RA, George J, Jowell P, Gress FG. Staging of esophageal cancer by EUS: staging accuracy revisited. *Gastrointest Endosc* 2007;**66**:475–82. http://dx.doi.org/10.1016/j.gie.2007.03.1051

101. Mennigen R, Tuebergen D, Koehler G, Sauerland C, Senninger N, Bruewer M. Endoscopic ultrasound with conventional probe and miniprobe in preoperative staging of esophageal cancer. *J Gastrointest Surg* 2008;**12**:256–62. http://dx.doi.org/10.1007/s11605-007-0300-2

102. Sandha GS, Severin D, Postema E, McEwan A, Stewart K. Is positron emission tomography useful in locoregional staging of esophageal cancer? Results of a multidisciplinary initiative comparing CT, positron emission tomography and EUS. *Gastrointest Endosc* 2008;**67**:402–9. http://dx.doi.org/10.1016/j.gie.2007.09.006

103. Takizawa K, Matsuda T, Kozu T, Eguchi T, Kato H, Nakanishi Y, *et al.* Lymph node staging in esophageal squamous cell carcinoma: a comparative study of endoscopic ultrasonography versus computed tomography. *J Gastroenterol Hepatol* 2009;**24**:1687–91. http://dx.doi.org/10.1111/j.1440-1746.2009.05927.x

104. 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**:547–57. http://dx.doi.org/10.1038/sj.bjc.6604200

105. Dyer SM, Levison DB, Chen RY, Lord SJ, Blamey S. Systematic review of the impact of endoscopic ultrasound on the management of patients with esophageal cancer. *Int J Technol Assess Health Care* 2008;**24**:25–35. http://dx.doi.org/10.1017/S026646230708004X

106. Harewood GC, Kumar KS. Assessment of clinical impact of endoscopic ultrasound on esophageal cancer. *J Gastroenterol Hepatol* 2004;**19**:433–9. http://dx.doi.org/10.1111/j.1440-1746.2003.03304.x

107. van Westreenen HL, Heeren PA, van Dullemen HM, van der Jagt EJ, Jager PL, Groen H, *et al.* Positron emission tomography with F-18-fluorodeoxyglucose in a combined staging strategy of esophageal cancer prevents unnecessary surgical explorations. *J Gastrointest Surg* 2005;**9**:54–61. http://dx.doi.org/10.1016/j.gassur.2004.09.055

108. Allum WH, Griffin SM, Watson A, Colin-Jones D. Guidelines for the management of oesophageal and gastric cancer. *Gut* 2002;**50**(Suppl. 5):1–23. http://dx.doi.org/10.1136/gut.50.suppl_5.v1

109. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967;**20**:637–48. http://dx.doi.org/10.1016/0021-9681(67)90041-0

110. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;**340**:c332. http://dx.doi.org/10.1136/bmj.c332

111. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:a1655.

112. North Wales Organisation for Randomised Trials in Health (NWORTH). *Standard operating procedures*. www.bangor.ac.uk/imscar/nworth/specservices.php?menu=3&catid-2236&subid=0 (accessed August 2011).

113. Fockens P, Van den Brande JH, van Dullemen HM, van Lanschot JJ, Tytgat GN. Endosonographic T-staging of esophageal carcinoma: a learning curve. *Gastrointest Endosc* 1996;**44**:58–62. http://dx.doi.org/10.1016/S0016-5107(96)70230-4

114. American Society of Anesthesiologists (ASA). *ASA Physical Status Classification System*. www.asahq.org/For-Members/Clinical-Information/ASA-Physical-Status-Classification-System.aspx (accessed August 2011).

115. World Health Organization (WHO). *WHO handbook for reporting results of cancer treatment.* Geneva: WHO; 1979.

116. Russell D, Hoare ZSJ, Whitaker R, Whitaker CJ, Russell IT. Generalised method for adaptive randomisation in clinical trials. *Stat Med* 2011;**30**:922–34.

117. Hoare Z. *Standard Operating Procedure for provision of randomisation service*. www.bangor.ac.uk/imscar/nworth/documents/5.01randomisationsopvs3_000.pdf (accessed August 2011).

118. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd edn. Hillsdale, NJ: Erlbaum; 1988.

119. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th edn. Oxford: Oxford University Press; 2008.

120. EuroQoL. *EQ-5D*. www.euroqol.org (accessed August 2011).

121. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;**37**:53–72. http://dx.doi.org/10.1016/0168-8510(96)00822-6

122. Dolan P. Modeling valuations for EuroQol health states. *Medical Care* 1997;**35**:1095–108. http://dx.doi.org/10.1097/00005650-199711000-00002

123. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, *et al.* The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;**11**:570–9.

124. Cella D. Assessment methods for quality of life in cancer patients: the FACIT measurement system. *Int J Pharmaceut Med* 2000;**14**:78–81. http://dx.doi.org/10.2165/00124363-200004000-00007

125. Eremenco S, Cashy J, Webster K. Linguistic validation of the FACT-Gastric (FACT-Ga) in Japanese and English. *Qual Life Res* 2003;**12**:818.

126. Darling G, Eton DT, Sulman J, Casson AG, Cella D. Validation of the functional assessment of cancer therapy esophageal cancer subscale. *Cancer* 2006;**107**:854–63. http://dx.doi.org/10.1002/cncr.22055

127. Hood K, Robling M, Ingledew D, Gillespie D, Greene G, Ivins R, *et al.* Mode of data elicitation, acquisition and response to surveys (MODE ARTS): systematic review. *Health Technol Assess* 2012;**16**(27).

128. Hahn EA, Cella D. Unbiased quality of life measurement across literacy levels and mode of administration. *Qual Life Res* 1997;**6**:654.

129. Glick HA, Doshi JA, Sonnad SS, Polsky D. *Economic evaluation in clinical trials*. Oxford: Oxford University Press; 2007.

130. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;**70**:213–20. http://dx.doi.org/10.1037/h0026256

131. Carpenter JR, Kenward MG. *Missing data in randomised controlled trials: a practical guide*. Southampton: National Evaluation Trials and Studies Coordinating Centre; 2007.

132. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing . . . presumed at random: cost-analysis of incomplete data. *Health Economics* 2003;**12**:377–92. http://dx.doi.org/10.1002/hec.766

133. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics* 1997;**53**:419–34. http://dx.doi.org/10.2307/2533947

134. Functional Assessment of Chronic Illness Therapy (FACIT). *Frequently asked questions*. www.facit. org/FACITorg/FAQ (accessed August 2011).

135. SPSS Inc. *SPSS 16.0 user's guide*. Chicago, IL: SPSS Inc.; 2007.

136. Winstead-Fry P, Schultz A. Psychometric analysis of the Functional Assessment of Cancer Therapy-General (FACT-G) scale in a rural sample. *Cancer* 1997;**79**:2446–52. http://dx.doi.org/10.1002/(SICI)1097-0142(19970615)79:12<2446::AID-CNCR23>3.0.CO;2-Q

137. SPSS Inc. *Amos 16.0 user's guide*. Chicago, IL: SPSS Inc.; 2007.

138. Kline RB. *Principles and practice of structural equation modeling*, 2nd edn. New York, NY: Guildford; 2005.

139. Jöreskog K, Sörbom D-S. *LISREL 8 user's reference guide*. Chicago, IL: Scientific Software International; 1996.

140. Williams J, Russell I, Durai D, Cheung W-Y, Farrin A, Bloor K, *et al.* What are the clinical outcome and cost-effectiveness of endoscopy undertaken by nurses when compared with doctors? Multi-Institution Nurse Endoscopy Trial (MINuET). *Health Technol Assess* 2006;**10**(40).

141. Edwards RT, Hounsome B, Linck P, Russell IT. Economic evaluation alongside pragmatic randomised trials: developing standard operating procedure for clinical trials unit. *Trials* 2008;**9**:64. http://dx.doi.org/10.1186/1745-6215-9-64

142. Drummond MF, Sculpher MJ, Torrance GW, O'Brien B, Stoddart GL. *Methods for the economic evaluation of health care programmes*. 3rd edn. Oxford: Oxford University Press; 2005.

143. Harris KM, Kelly S, Berry E, Hutton J, Roderick P, Cullingworth J, *et al.* Systematic review of endoscopic ultrasound in gastro-oesophageal cancer. *Health Technol Assess* 1998;**2**(18).

144. Lennon AM, Penman ID. Endoscopic ultrasound in cancer staging. *Br Med Bull* 2007;**84**:81–98. http://dx.doi.org/10.1093/bmb/ldm033

145. Hadzijahic N, Wallace M, Hawes R, VanVelse A, LeVeen M, Marsi V, *et al.* CT or EUS for the initial staging of esophageal cancer? Cost minimization analysis. *Gastrointest Endosc* 2000;**52**:715–20. http://dx.doi.org/10.1067/mge.2000.108481

146. Harewood GC, Wiersema MJ. A cost analysis of endoscopic ultrasound in the evaluation of esophageal cancer. *Am J Gastroenterol* 2002;**97**:452–8. http://dx.doi.org/10.1111/j.1572-0241.2002.05499.x

147. Wallace MB, Nietert PJ, Earle C, Krasna MJ, Hawes RH, Hoffman BJ, *et al.* Analysis of multiple staging management strategies for carcinoma of the esophagus: computed tomography, endoscopic ultrasound, positron emission tomography and thoracoscopy/laparoscopy. *Ann Thorac Surg* 2002;**74**:1026–32. http://dx.doi.org/10.1016/S0003-4975(02)03875-4

148. National Institute for Health and Care Excellence (NICE). *Guide to the methods of technology appraisal (consultation document)*. www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (accessed August 2011).

149. Brazier J, Ratcliffe J, Salomon JAT. *Measuring and valuing health benefits for economic evaluation*. 1st edn. Oxford: Oxford University Press; 2007.

150. Knapp M, Beecham J. Costing mental health services. *Psychol Med* 1990;**20**:893–908. http://dx.doi.org/10.1017/S003329170003659X

151. Ridyard CH, Hughes DA. Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK Health Technology Assessment Programme. *Value Health* 2010;**13**:867–72. http://dx.doi.org/10.1111/j.1524-4733.2010.00788.x

152. Department of Health (DoH). *NHS reference costs 2007–08*. London: DoH; 2009.

153. NHS Information Centre. *Prescription Cost Analysis 2008 England*. www.ic.nhs.uk/statistics-and-data-collections/primary-care/prescriptions/prescription-cost-analysis-2008 (accessed August 2011).

154. Shenfine J, McNamee P, Steen N, Bond J, Griffin SM. A pragmatic randomised controlled trial of the cost-effectiveness of palliative therapies for patients with inoperable oesophageal cancer. *Health Technol Assess* 2005;**9**(5).

155. Curtis L. *Unit costs of health and social care 2008*. Canterbury: PSSRU, University of Kent; 2008.

156. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves: facts, fallacies and frequently asked questions. *Health Econ* 2004;**13**:405–15. http://dx.doi.org/10.1002/hec.903

157. Fayers PM, Hand DJ, Bjordal K, Groenvold M. Causal indicators in quality of life research. *Qual Life Res* 1997;**6**:393–406. http://dx.doi.org/10.1023/A:1018491512095

158. North Wales Organisation for Randomised Trials in Health (NWORTH). Standard operating procedure for safety monitoring including pharmacovigilance. Bangor: Bangor University; 2011. www.bangor.ac.uk/imscar/nworth/documents/4.03safetymonitoringincpharmacovigilancevs5.pdf (accessed September 2012).

159. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011;**342**:d40. http://dx.doi.org/10.1136/bmj.d40

160. Beseth BD, Bedford R, Isacoff WH, Holmes EC, Cameron RB. Endoscopic ultrasound does not accurately assess pathologic stage of esophageal cancer after neoadjuvant chemoradiotherapy. *Am Surg* 2000;**66**:827–31.

161. Gierada DS, Garg K, Nath H, Strollo DC, Fagerstrom RM, Ford MB. CT quality assurance in the lung screening study component of the National Lung Screening Trial: implications for multicenter imaging trials. *Am J Roentgenol* 2009;**193**:419–24. http://dx.doi.org/10.2214/AJR.08.1995

162. Toita T, Oguchi M, Ohno T, Kato S, Niibe Y, Kodaira T, *et al.* Quality assurance in the prospective multi-institutional trial on definitive radiotherapy using high-dose-rate intracavitary brachytherapy for uterine cervical cancer: the individual case review. *Jpn J Clin Oncol* 2009;**39**:813–19. http://dx.doi.org/10.1093/jjco/hyp105

163. Velasquez L, Boellaard R, Kollia G, Hayes W, Hoekstra O, Lammertsma A, *et al.* Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. *J Nucl Med* 2009;**50**:1646–54. http://dx.doi.org/10.2967/jnumed.109.063347

164. Fockens P, Kisman K, Merkus MP, van Lanschot JJ, Obertop H, Tytgat GN. The prognosis of esophageal carcinoma staged irresectable (T4) by endosonography. *J Am Coll Surg* 1998;**186**:17–23. http://dx.doi.org/10.1016/S1072-7515(97)00131-2

165. Chak A, Cento M, Gerdes H, Kightdale C, Hawes R, Wiersema M, *et al.* Prognosis of esophageal cancers preoperatively staged to be locally invasive (T4) by endoscopic ultrasound (EUS): a multicenter retrospective cohort study. *Gastrointest Endosc* 1995;**42**:501–6. http://dx.doi.org/10.1016/S0016-5107(95)70001-3

166. Waxman I, Raju GS, Critchlow J, Antoniolli DA, Spechler SA. High frequency probe ultrasonography has limited accuracy for detecting invasive adenocarcinoma in patients with Barrett's esophagus and high grade dysplasia or intramucosal cancer; a case series. *Am J Gastroenterol* 2006;**101**:1773–9. http://dx.doi.org/10.1111/j.1572-0241.2006.00617.x

167. Pech O, May A, Günter E, Gossner L, Ell C. The impact of endoscopic ultrasound and computed tomography on the TNM stage of early cancer in Barrett's esophagus. *Am J Gastroenterol* 2006;**101**:2223–9. http://dx.doi.org/10.1111/j.1572-0241.2006.00718.x

168. Claxton K, Walker S, Palmer S. *Appropriate perspectives for health care decisions*. York: University of York Centre for Health Economics; 2010.

169. Hughes DA, Tunnage B, Yeo ST. Drugs for exceptionally rare diseases: do they deserve special status for funding? *Q J Med* 2005;**98**:829–36. http://dx.doi.org/10.1093/qjmed/hci128

Appendix 1 Trial protocol

Project title

Cancer of Oesophagus or Gastricus: New Assessment of Technology of Endosonography (COGNATE).

Funding

National Coordinating Centre for Health Technology Assessment (NCCHTA) [project number 01/01/03].

Planned investigation

Research objectives

What is the problem to be addressed?
The aim of this trial is to evaluate the role that endoscopic ultrasound (EUS) staging plays in the management of patients with gastro-oesophageal cancer (GOC). Specific objectives:

1. To estimate the marginal effect of EUS staging, compared with a standard staging algorithm, on the selection of treatment for patients with GOC viz the proportions of patients treated surgically, with multimodality treatment, or without surgery.
2. To estimate the effect of EUS staging on the outcome of care of patients with GOC. The outcome indicators include:
   a. Primary outcome: Quality-adjusted survival
   b. Secondary outcomes:
      i.   Proportion of patients undergoing a complete resection.
      ii. Survival.
      iii. Quality of survival.
      iv. Resource use including treatment and subsequent use of health care.
3. To assess the cost-effectiveness of EUS by comparing improvements in patient outcomes with the marginal cost of EUS.
4. To estimate the proportion of patients with GOC who benefit from EUS and thus to model the need for EUS facilities within a defined population.

How the results of this trial will be used
The results of this multi-centre trial will inform policy whether EUS staging should be routine for patients with GOC and for which patients it is likely to benefit outcome. These data will be useful in modelling the resources required for EUS within a defined population.

The trial will provide robust evidence to underpin the development of guidelines for the effective staging of patients with GOC. We shall circulate such guidelines, initially among participating centres, but thereafter through national groups.

Existing research

Background
The Scottish Audit of Gastric and Oesophageal Cancer (SAGOC) is a prospective audit of gastric and oesophageal cancers. Data are available on 3300 patients with gastric and oesophageal cancers, that is

more than 95% of those diagnosed during 1997–1999. The data from SAGOC provide a population-based description of the current treatment of GOC and also predict the numbers of patients likely to be eligible for trials or treatments.

The change in the demography of upper gastro-intestinal cancers, well documented in Western series, is characterised by an increased incidence of lower oesophageal and proximal gastric cancers but a decreased incidence of distal gastric cancers.[1–3] Figures from Scottish and other UK centres have reflected the trend elsewhere – an increase in oesophageal cancers of 2% a year.[4,5] Unfortunately the survival of these patients is poor (*Table 1*), and treatment is associated with significant morbidity and mortality.[6–10]

The prognosis of patients with GOC depends on pre-morbid status, as many of these cancers occur in elderly and frail patients. Grading systems like that of the American Society of Anaesthesiologists (ASA) can predict survival (*Figure 1*). The World Health Organisation (WHO) performance status of patients can also predict survival. In patients with a good pre-morbid status, prognosis depends on the stage of the tumour; patients with metastatic disease have a poor prognosis.[11,12] For those with localised disease the anatomical extent of the tumour and nodal status are the most important prognostic indicators[7,13] (*Figure 2*).

In patients in whom surgery was performed SAGOC identified the following independent predictors of one-year survival: ASA grading; curative versus palliative intent of surgery; incomplete resections; and complications associated with surgery.

The most commonly used staging techniques were trans-abdominal ultrasonography scan (USS), chest X-ray (CXR), contrast-enhanced computerised tomography (CT) scan, and laparoscopy in selected patients; there was little use of magnetic resonance imaging (MRI) at that time (*Table 2*). Similar results were reported from Wales.[14] When centres use only these conventional staging techniques many patients undergo non-curative resections.[7,9] SAGOC has shown under-staging of oesophageal and gastro-oesophageal junctional cancers in approximately 20% of patients with operable tumours. This results in non-curative resections with macroscopic or microscopic tumour remaining, and does not improve length or quality of survival.[15] This is a particular problem for patients with T3 tumours, i.e. those that have breached the wall of the oesophagus or stomach but are not invading adjacent structures. Alternatives to surgery using chemotherapy and radiotherapy appear to be effective in such patients with advanced localised tumours.[16–19]

Alternative treatments are also now available for patients with cancers limited to the mucosa. Such tumours have a low risk of lymph node metastases and may be treated with mucosal ablative techniques such as endoscopic mucosal resection (EMR). Such treatment has been shown to be effective for mucosal cancers and avoids both immediate and long-term problems associated with surgery.[20,21] As tumours invade the sub-mucosal layer, however, there is increased risk of lymphatic dissemination and EMR is not an adequate treatment.[22] Using standard staging techniques it is difficult to differentiate between mucosal and sub-mucosal tumours and thus select appropriate patients for EMR.

Endoscopic ultrasound and staging of GOC

Initial reviews of CT, MRI and trans-abdominal ultrasonography suggested a tendency to under-stage tumours; furthermore MRI does not appear to improve good quality CT.[23–26] New developments may improve our ability to stage gastric and oesophageal cancers. The introduction of modern CT protocols and spiral images have improved the accuracy of CT staging.[27] Endoscopic ultrasonography (EUS) allows a high frequency ultrasound probe mounted at the end of an endoscope to be placed directly against either an oesophageal or gastric cancer. The high frequency of the ultrasound improves the spatial resolution that can be obtained and may thus improve the accuracy of tumour staging.[28–30] The systematic review preceding COGNATE showed that relative to conventional techniques EUS improves the accuracy of staging local GOC and the nodal status of these tumours[31] (*Table 3*).

TABLE 1 Survival from diagnosis of gastro-oesophageal cancer patients in Scotland (SAGOC data)

Tumour	6 months	12 months	18 months	24 months	30 months
Oesophagus	51.4%	29.0%	19.8%	13.8%	11.1%
OG junction	49.6%	32.7%	23.1%	16.6%	10.7%
Gastricus	50.7%	34.5%	25.7%	20.4%	17.4%
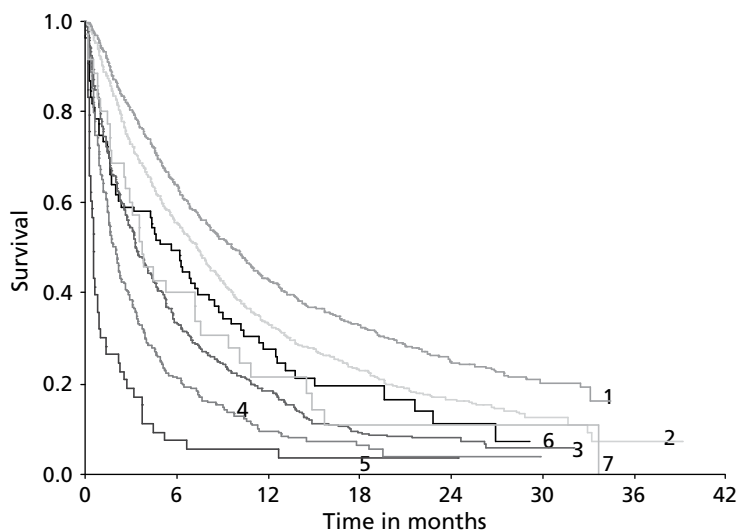
FIGURE 1 Unadjusted survival from localised malignancy (Kaplan–Meier estimates) by ASA grading (SAGOC data). **Key** (ASA 1) no physiological disturbance (ASA 2) minor physiological impairment (ASA 3) significant physiological impairment but responsive to treatment with medication (ASA 4) severe physiological impairment not responsive to treatment (ASA 5) patient moribund (6) grade not recorded (7) average survival of SAGOC patients.
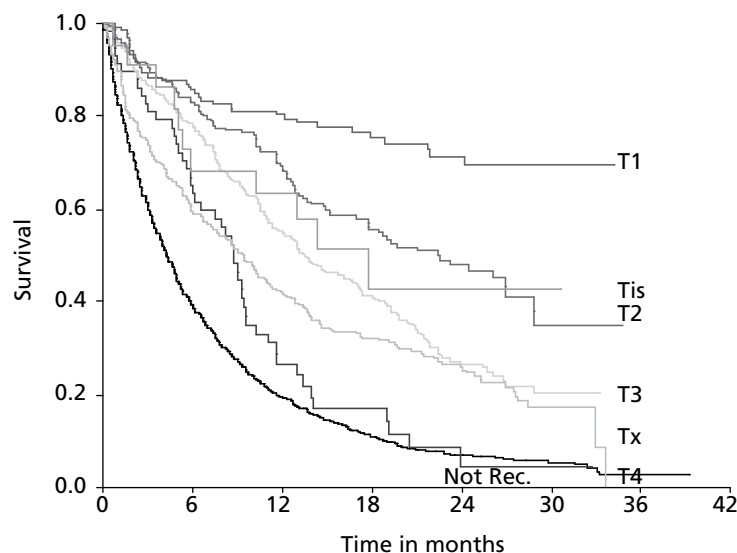
FIGURE 2 Unadjusted survival (Kaplan–Meier estimates) by stage (SAGOC data). **Key** (Tx) tumour not evaluable (Tis) carcinoma in-situ (T1) tumour limited to mucosa or sub-mucosa (T2) tumour invades muscularis propria (T3) tumour invades adventitia (T4) tumour involves adjacent structures (Not Rec) Stage not recorded.

TABLE 2 Staging investigations used in population based studies: percentage of patients with given tumour who received specified investigation

	SAGOC			Wales[18]	
	Oesophageal	**OG junction**	**Gastric**	**Oesophageal**	**Gastric**
CT	71%	73%	60%	70%	59%
CXR	62%	61%	65%	NR	NR
USS	22%	30%	40%	64%	58%
Laparoscopy	16%	26%	23%	5%	16%
Bronchoscopy	7%	2%	1%	–	–
Endoscopic US	4%	4%	2%	–	–
MRI	1%	0.4%	0.5%	–	–

TABLE 3 Reported sensitivity and specificity of endoscopic ultrasonography[31]

Category	Sensitivity range	Specificity range
Oesophageal (T)	71–100%	67–100%
Gastro-oesophageal junction (T)	42–100%	67–100%
Gastric (T)	68–100%	88–100%
Nodes (N)	60–97%	40–100%

The need for a trial

Though EUS has been recommended as essential in staging of oesophageal cancers,[32] it has not been critically assessed. There are problems in EUS staging of non-traversable tumours;[33] the majority of these are T3 or T4 lesions, which need better staging to avoid non-curative resections. EUS was least accurate in carcinomas around the gastro-oesophageal junction[33] – the tumours which are increasing most rapidly in incidence. Furthermore there are few studies comparing the value of EUS with that of modern CT protocols.[33] Therefore, although there is evidence that EUS improves the anatomical staging of GOC, it is not clear how it affects patient management. In particular it is not clear whether there is any benefit from adding EUS to contemporary staging protocols based on helical CT techniques. SAGOC shows that before COGNATE only a minority of patients with GOC underwent EUS (*Table 2*). Hence the decision to use EUS more widely depends crucially on whether evaluation shows that it is of major benefit to patients. Therefore it is essential to estimate the effect of EUS staging on the management of gastric and oesophageal cancers. Furthermore it is important to know the proportion of patients with GOC that are likely to benefit from EUS. To answer these questions rigorously needs a randomised controlled trial which assesses patients by a conventional staging algorithm and then randomises them between EUS or not.

EUS may especially benefit three groups of patients with GOC:

1. Patients with T1 tumours localised to the mucosa, which EUS may identify as likely to benefit from endoscopic treatment, thus avoiding unnecessary surgery.
2. Patients with tumours which EUS may discriminate as either likely to benefit from 'curative' surgery' or to have residual disease after major surgery with attendant risks.
3. Patients with T3 or T4 tumours which EUS may identify as likely to benefit from multi-modal treatment or not.

Research design

Summary

Participating centres will use a defined staging algorithm based on usual practice for patients with GOC. Patients with localised tumours will be randomised to receive EUS or not after stratification by location – gastric, oesophageal or gastro-oesophageal junction. In both groups multi-disciplinary teams will choose between three main treatments:

1. Tumours adjudged mucosal will undergo endoscopic mucosal resection (EMR) with or without argon beam ablation of the surrounding mucosa.
2. Tumours adjudged resectable will undergo surgical resection, with or without neo-adjuvant cisplatin and 5FU.
3. Tumours adjudged not to be resectable will receive chemotherapy, with or without radiotherapy depending on the site.

We shall compare the two groups for treatment received, rate of complete resections, and length and quality of survival.

Trial design

Figures 3 and *4* summarise the trial design. A multi-centre randomised controlled trial will ensure that there is no bias in the selection of patients for EUS or not. Randomisation will take place after the initial staging investigations have been completed and reviewed at a multi-disciplinary meeting. At this stage clinicians will agree a conditional management plan and randomise patients either to receive EUS or to proceed directly to the agreed management plan. They will report patients whom they decide not to randomise to the trial co-ordinating centre in Bangor, with the reasons for their exclusion.

Interventions

We developed the staging algorithm from usual practice, as identified by SAGOC before the trial began:

a. Chest x-ray, pulmonary function tests, haematology and biochemistry, together with assessment of cardiac status. Patients of WHO performance status 3 or 4 or medically unsuitable for either surgery or chemotherapy will be excluded.
b. Patients who are medically fit will undergo a trans-abdominal USS. Those found to have metastatic liver disease will be excluded.
c. Patients without evidence of metastases will undergo a CT scan following an agreed protocol using a spiral scanner, oral water contrast and intravenous contrast. Laparoscopy will be undertaken in patients with any suspicion of peritoneal disease, as this remains the best means of detecting peritoneal deposits of tumour.[11]
d. Only patients with localised tumours will be randomised between EUS or not.

In the resulting non-EUS group the choice of treatment will depend on the results of these standard investigations. In the EUS group that choice will follow the extra investigation. At the end of staging, with or without EUS, multi-disciplinary teams will allocate patients to one of three treatment groups:

1. Patients adjudged to have mucosal tumours will be treated with EMR and the surrounding mucosa ablated.
2. Patients with tumours adjudged to be resectable will be treated with surgery, with or without neo-adjuvant chemotherapy.
3. Patients with advanced localised disease for which a complete resection is not adjudged possible will receive multi-modal treatment, possibly including palliative surgery in patients with gastric cancers.
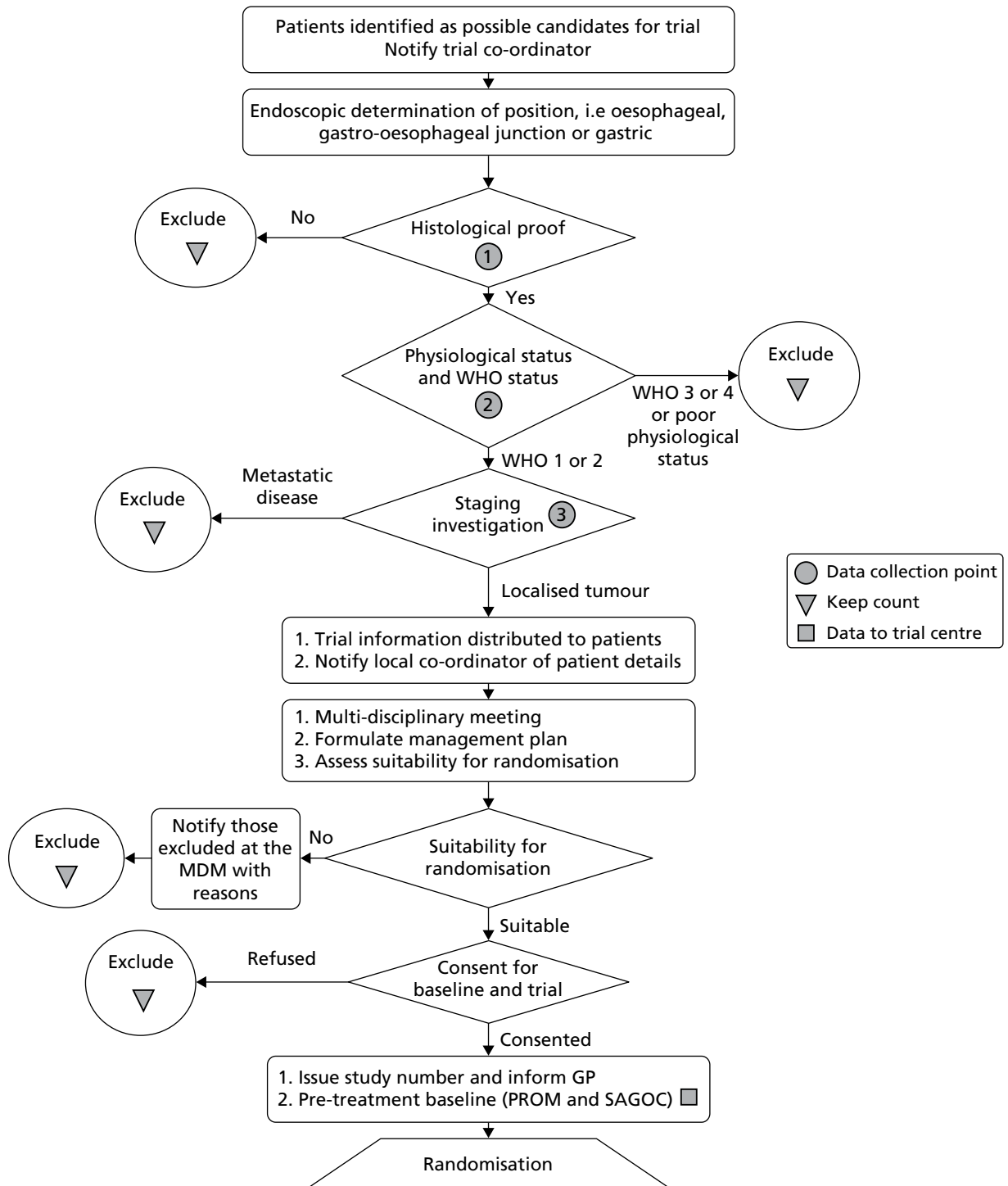
FIGURE 3 Trial design: start to randomisation.
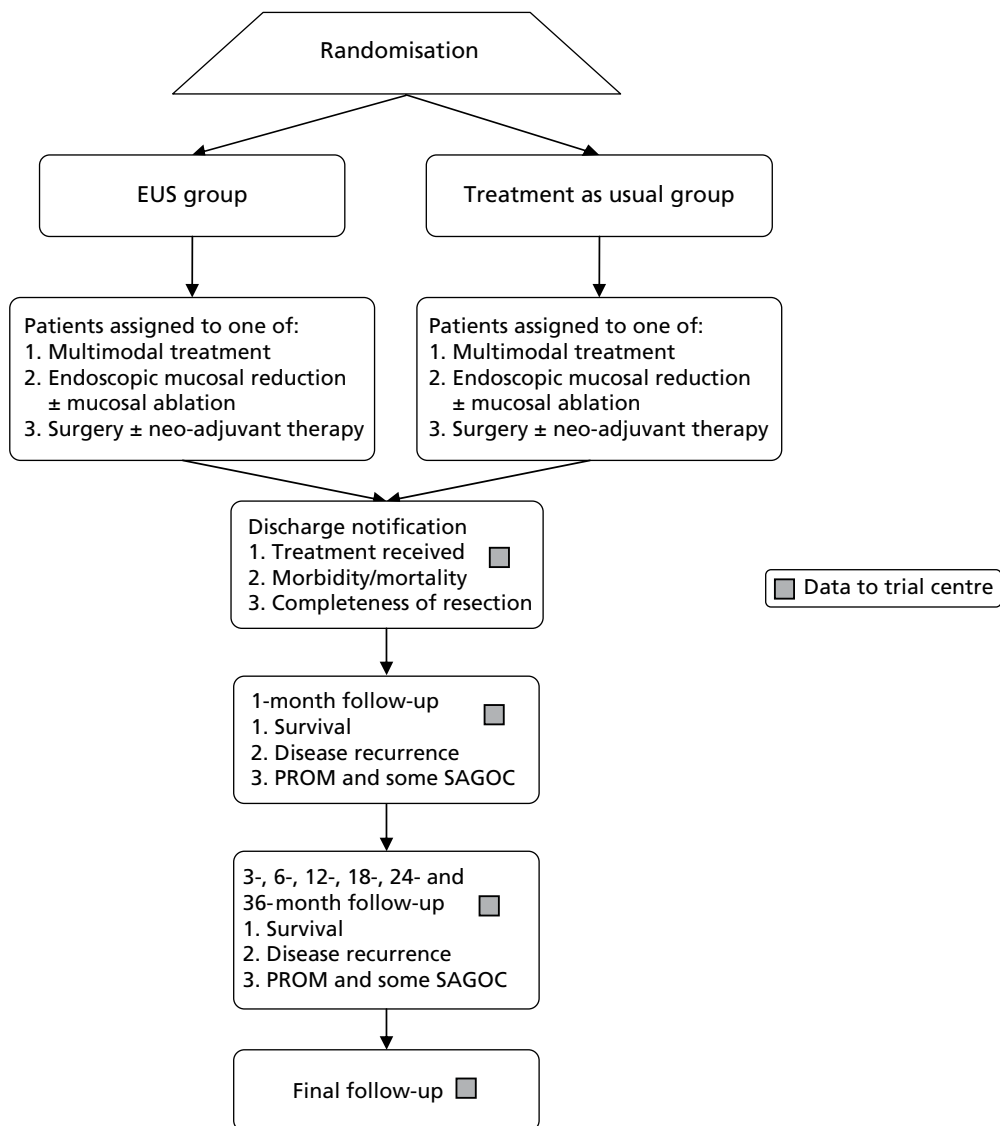
FIGURE 4 Trial design: randomisation to conclusion.

Inclusion and exclusion criteria

To be randomised patients should be fit for both surgery and chemo-radiotherapy and free of metastatic disease. Their ASA grade should be 1 or 2 (*Figure 1*) and their WHO performance status also 1 or 2. Following initial staging clinicians will identify all eligible patients but may exclude patients from the trial for clinical reasons. To ensure there is no bias we shall monitor all such exclusions and the corresponding reasons.

Randomisation

We shall randomise consented eligible patients by telephone to the COGNATE office in Bangor after stratification for centre and tumour location viz gastric, oesophageal, or at the gastro-oesophageal junction. As only patients with a good performance status will be randomised there is no need to stratify for performance status.

Follow up

Median life expectancy with GOC is 18 months. We shall follow patients until death or the end of data collection 12 months after the end of recruitment. We shall collect data at the time of discharge from hospital after initial treatment and thereafter every three months.

Ethical considerations

The value of EUS in the staging of patients with GOC is not proven. The only ethical means of evaluating this investigation is therefore a randomised controlled trial.

Risks and benefits

The technique of EUS is safe and carries the same risk to the patient as an endoscopy. It has the potential to improve staging of gastric and oesophageal tumours, particularly the T stage, and thus provide prognostic information which may guide management. However it is not clear how staging influences management or whether EUS improves management decisions. Accordingly there is potential for patients to be assigned to a management plan which may disadvantage them. Therefore we shall carefully monitor the quality of treatment under the surveillance of the COGNATE Data Monitoring and Ethics Committee (DMEC), with the intention of detecting increased morbidity or mortality in any treatment or any centre at an early stage. If EUS can improve the selection of treatment, this will benefit, not only individual patients, but also the population of patients with GOC, as it will encourage better targeting of resources. However if COGNATE shows that EUS does not improve selection, the number of procedures will fall, releasing resources for GOC and other patients.

Patient information and informed consent

Before randomisation we shall ask eligible patients with GOC to participate in the COGNATE trial. We shall explain the process of randomisation to patients, together with the nature of EUS. We shall stress that the subsequent choice of treatment is identical in both groups.

Data storage

All data will identify each patient only by a unique trial number. Each trial centre will keep its own index linking trial numbers to patients' names and addresses separate from the laptop computers used to store and transfer trial data, and protect that index by key and password. Those in Bangor analysing the data will have no access to these local indices.

Sample size

We aim to consent, randomise and follow up a total of 400 patients. As there is no easy means of calculating the power of this sample for the primary outcome of quality-adjusted survival, we calculate power for two simple but plausible scenarios. First if we assume no difference between groups in quality of life (as measured for example by FACT-GE) a log-rank test using a 5% significance level would yield 80% power of detecting a hazard ratio of 0.6, equivalent to a difference between 60% and 73% in survival at 12 months (derived from SAGOC data by analogy with *Figure 2*). Secondly if we assume no difference between groups in survival, a t test using a 5% significance level would yield greater than 80% power of detecting a 'small' effect size of 0.3 in quality of life. As the groups are more likely to differ in both survival and quality, the power of our primary analysis of quality-adjusted survival will be correspondingly greater. At worst to consent, randomise and follow up only 220 patients will yield 80% power to detect a hazard ratio of 0.5 (equivalent to a difference between 60% and 78% in survival at 12 months) or an effect size of 0.4 in quality of life, still a 'small' effect.

Outcome measures

We shall use the following measures to compare the two randomised groups:

1. *Primary outcome measure:* Quality adjusted survival
   We shall ask patients to attend follow-up clinics after 1, 3, 6, 12, 18, 24 and 36 months and assess their quality of life through the EuroQol EQ-5D and the Functional Assessment of Cancer Therapy

(FACT), in particular the general module FACT-G,[34] the oesophageal module FACT-E and the gastric module FACT-Ga. As FACT-E and FACT-Ga have many questions that are common or similar, the FACT team have permitted and encouraged us to combine them into a gastro-oesophageal module provisionally called FACT-GE. Two of us with substantial experience of validating patient-assessed outcome measures (ITR & DKI) are concurrently revalidating FACT-GE using the methods described by Streiner & Norman.[35] At worst we shall adjust survival by the EQ-5D in traditional fashion. If FACT-GE proves more responsive to change than EQ-5D, as we expect, we shall adjust survival by FACT-GE. We shall therefore complete our concurrent validation of FACT-GE and finalise the COGNATE analysis plan before starting the definitive analysis.

2. *Secondary outcome measures*:
   a. Quality of treatment:
      i. Complete resection rate. This will include pathological data on both EMR and resected tumours. For patients treated with EMR we shall record residual tumours and any additional treatment.
      ii. Pathological reporting of resected specimens according to the SAGOC recommendations,[36] under the surveillance of the DMEC.
      iii. Treatment-related morbidity and mortality according to the SAGOC definitions.[36] In particular mortality will include deaths in hospital following treatment or within 30 days of treatment.
   b. Survival – to 12 months for those last randomised and to 48 months for those first randomised.
   c. Quality of survival – FACT-G and FACT-GE.

3. *Health economics*:
   Within COGNATE we shall assess whether EUS is more cost-effective than conventional staging in the diagnosis and treatment of GOC by estimating the incremental cost-effectiveness ratio of EUS relative to conventional staging. We shall estimate differences in the cost of patients' care between the two groups and relate this to differences in effectiveness in the form of quality adjusted survival. Following COGNATE we shall use SAGOC data on the prognosis of patients with GOC to model the long-term costs and benefits of EUS.
   a. Measurement of effectiveness. For the purpose of estimating the cost-effectiveness ratio, we shall measure effectiveness in Quality Adjusted Life-Years (QALYs). Nevertheless we shall set this calculation within a broader cost-consequence analysis that will include the full range of primary and secondary outcomes of the trial, for example generating both cost per life-year gained and cost per QALY gained.
   b. Measurement of costs. We shall analyse COGNATE from the perspective of the NHS, covering the major direct costs of health care resources used by patients in the trial. These costs will include initial treatment and subsequent investigation, treatment and palliation, and other major elements of primary and secondary care. The local co-ordinators at each of the sites uses an electronic database to record the main uses of NHS resources by trial patients throughout the study period. We shall also ask NHS finance departments at each site to provide unit costs for procedures received by patients in the trial. Finally we shall compare the putative costs of the treatment plans proposed after initial staging in each group with that adopted following EUS in the experimental group, and the actual costs of treatment in each group.
   c. Sensitivity and threshold analysis. We shall conduct sensitivity and threshold analyses based on the observed distributions of outcomes and costs, to test whether, and to what extent, the incremental cost-effectiveness ratio of EUS relative to conventional staging is sensitive to key assumptions in our analysis. We shall use 'bootstrapping' to estimate skewed costs in unbiased fashion and cost effectiveness acceptability curves to interpret findings.
   d. Generalisability and policy implications. We shall compare the findings of this economic evaluation, in particular the estimated cost per QALY, including confidence intervals, with those available from other studies at the end of our trial. This approach will enable us to place the cost-effectiveness of the diagnosis and treatment of GOC, both with and without EUS, within the range of estimated health gains 'per NHS pound' for other conditions.

Statistical analysis

Primary analysis will be by "intention to investigate by EUS". This reflects the essentially pragmatic nature of the trial, and its primary goal of assessing health technology to inform decisions in the real world. We shall also undertake secondary analysis by "EUS received" to explore the implications of documented clinical decisions to diverge from the allocated algorithm. The primary survival analysis will use site and stage as covariates, especially if there is any evidence of baseline imbalance between groups despite stratification and remote randomisation. We shall analyse secondary outcome measures by general linear models, again allowing us to use covariates, notably site and stage, when appropriate.

Trial management

The COGNATE Trial has a Steering Committee comprising an independent chair, two independent members and four members of the COGNATE trial executive group (TEG). Lead clinicians in each centre report to the Trial Steering Committee (TSC) through the TEG. The independent DMEC also reports to the TSC (*Figure 5*).

Composition of Trial Steering Committee

Robert Heading, Consultant Gastroenterologist, Royal Infirmary, Edinburgh (chair).
Hugh Gilmour, Consultant Pathologist, University of Edinburgh (independent member).
Toni Lerut, Professor of Thoracic Surgery, University of Leuven (international member).
COGNATE team – Kenneth Park, Ian Russell, Grant Fullarton, Shona Campbell.

Composition of Data Monitoring and Ethics Committee

Hugh Gilmour, Consultant Pathologist, University of Edinburgh (chair).
Marion Campbell, Statistician, Health Services Research Unit, University of Aberdeen.
David Kirby, Chair of Oesophageal Cancer Patients Group.

Project timetable and milestones

Timetable

Our initial plan was to conduct COGNATE over a period of 60 months from February 2004 through January 2009. However a combination of factors beyond the control of the co-ordinating centre (discussed with the visiting party from NCCHTA in April 2006) has slowed both recruitment and expenditure. In these circumstances we are actively considering applying for a 'no-cost extension' through October 2009. The following timetable displays the current timetable without brackets and the potential extension in brackets:

1. Start of funding: 01/02/04
2. Detailed design & pilot study to ensure consistency of staging investigations & treatment options, and that randomisation and data collection are robust.
3. Start of recruitment: 01/02/05
4. End of recruitment: 31/07/08
5. End of follow-up to ensure minimum of 12 months for all trial participants: 31/07/09
6. Analysis and draft report covering all patients – 31/01/10
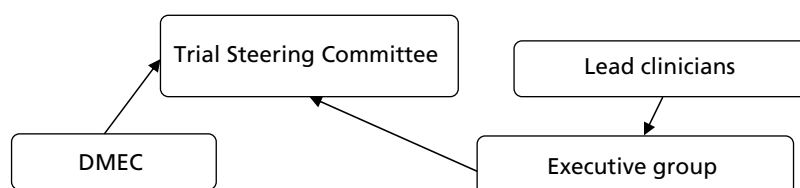
FIGURE 5 Trial management.

Recruitment

Over the two years 2005 and 2006 eight centres joined the original six centres. On closing the recruitment of centres at the end of 2006 we revised our power calculations and recruitment targets in the light of the numbers of patients with GOC seen in the original centres over the first 23 months of recruitment. From SAGOC we estimate that one third will be eligible for inclusion. From our feasibility study of recruitment we estimate that 90% of these will consent to randomisation.

Expertise and responsibilities of applicants

- Prof Ken Park (KGMP) is consultant surgeon in the upper gastrointestinal surgical unit at Aberdeen Royal Infirmary. He is co-chair of the Scottish Audit of Gastro-Oesophageal Cancer (SAGOC). He is responsible for the clinical management of COGNATE. He also acts as lead clinician in Aberdeen. As co-chief investigator with ITR, he sits on the Trial Steering Committee.
- Prof Ian Russell (ITR) specialises in the design, conduct and analysis of pragmatic randomised trials. Since October 2002 he been Director of the Institute of Medical and Social Care Research (IMSCaR) at University of Wales Bangor, which includes centres devoted to clinical trials, the economics of health, public health, and social care. IMSCaR enjoys a close relationship, both organisational and geographical, with the University Department of Psychology (RAE rating 5*), which has particular strengths in clinical and health psychology. ITR is responsible for the technical management of COGNATE and for supervising the trial team in Bangor. As co-chief investigator with KGMP, he sits on the Trial Steering Committee.
- Mr Stephen Attwood (SEAA) is consultant surgeon with the Northumberland NHS Trust with a particular interest in gastro-intestinal cancer. He has written extensively on endoscopic techniques and has pioneered endoscopic ablative treatment. He advises the trial executive group.
- Prof Hugh Barr (HB) is consultant surgeon at Gloucester Royal Infirmary and Dean of the Postgraduate Medical School of Cranfield University. He has an international reputation for treatment of gastric and oesophageal cancer. He is lead clinician in Gloucester.
- Dr Shona Campbell (SC) is consultant radiologist at the University Hospital of Leicester. She has a longstanding interest in endoscopic ultrasonography. She is lead clinician in Leicester and sits on the Trial Steering Committee.
- Dr Rhiannon Edwards (RTE) is Director of the Centre for the Economics of Health within IMSCaR at the University of Wales Bangor. She specialises in economic evaluation and modelling. She leads the economic evaluation of EUS in COGNATE. She will use SAGOC data to develop a broader model of the implications of COGNATE findings for the effectiveness of EUS across the NHS.
- Mr Grant Fullarton (GMF) is consultant surgeon with North Glasgow University NHS Trust. He has a particular interest in upper gastro-intestinal surgery and endoscopic techniques. He acts as lead clinician in Glasgow. As the representative of the largest centre he sits on the Trial Steering Committee.
- Prof Fiona Gilbert (FJG) is an academic radiologist in the Department of Radiology, University of Aberdeen, with expertise in the design and management of trials in radiology. She is co-chair of SAGOC. She has a particular interest in the quality of pre-operative staging investigations, and will monitor the quality of these investigations across COGNATE.
- Dr David Ingledew (DKI) is Senior Lecturer in Psychology at the University of Wales Bangor. He is a health psychologist and psychometrician with expertise in the development and validation of health-related measurement scales. He will contribute to the development, validation and analysis of patient-assessed outcome measures, including quality of life.

References

1. Blot W. Esophageal cancer trends and risk factors. *Sem Oncol* 1994;**21:**403–10.

2. Hansson L, Sparen P, Nyren O. Increasing incidence of both major histological types of esophageal carcinomas amongst men in Sweden. *Int J Cancer* 1993;**54:**402–7.

3. Devesa S, Blot W, Fraumeni J. Changing patterns in the incidence of esophageal an gastric carcinoma in the United States. *Cancer* 1998;**83:**2049–53.

4. McKinney P, Sharp L, McFarlane G. *Oesophageal and Gastric cancer in Scotland* 1960–1990. *Br J Cancer* 1995;**71:**411–15.

5. Office for National Statistics. *Registration of cancer diagnosed in* 1993–1996, England and Wales. *Health Stat Q* 1999;**4:**59–70.

6. Allum WH, Powell DJ, McConkey CC, Fielding JWL. Gastric cancer a 25 year review. *Br J Surg* 1989;**76:**535–40.

7. Muller JM, Erasmit T, Stelsner M, Zieren U, Pichmaier H. Surgical therapy of oesophageal cancer. *Br J Surg* 1990;**77:**845–57.

8. MacIntyre IMC, Akoh JA. Improving survival in gastric cancer. *Review of operative mortality in English language publications from* 1970. *Br J Surg* 1991;**78:**773–8.

9. Cushieri A, Fayers P, Fielding J, *et al.* Postoperative morbidity and mortality after D1 and D2 resections for gastric cancer: preliminary results of the MRC randomised controlled surgical trial. *Lancet* 1996;**347:**995–9.

10. Bonnenkamp JJ, Songun J, Hermans J, *et al.* Randomised comparison of morbidity after D1 and D2 dissection for gastric cancer in 996 Dutch patients. *Lancet* 1995;**345:**745–8.

11. Anderson DN, Campbell S, Park KGM. The accuracy of laparoscopic ultrasound in the staging of upper gastrointestinal malignancy. *Br J Surg* 1996;**83:**1424–8.

12. Oliver SE, Robertson CS, Logan RFA. Oesophageal cancer: a population based study of survival after treatment. *Br J Surg* 1992;**79:**1321–5.

13. Nakamura K, Ueyama T, Yao T, *et al.* Pathology and prognosis of gastric carcinoma. Findings in 10,000 patients who underwent primary gastrectomy. *Cancer* 1992;**70:**1030.

14. Pye JK, Crumplin MKH, Charles J, Kerwat R, Foster ME, Biffin A. One-year survey of carcinoma of the oesophagus and stomach in Wales. *Br J Surg* 2001;**88:**278–85.

15. De Boer AGEM, Onorbe Genovesi PI, Sprangers MAG, van Sandick JW, Obertop H, van Lanschot JJB. Quality of life in long-term survivors after curative transhiatal oesophageactomy for oesophageal carcinoma. *Br J Surg* 2000;**87:**1716–21.

16. Cooper JS, Guo MD, Herskovic A, *et al.* Chemoradiotherapy of locally advanced esophageal cancer: long-term follow-up of a prospective randomised trial (RTOG 85–01). Radiation Therapy Oncology Group. *JAMA* 1999;**281:**1623–7.

17. Walsh TN, Noonan N, Hollywood D, *et al.* A comparison of multimodality therapy and surgery for esophageal adenocarcinoma. *N Engl J Med* 1996;**335:**462–7.

18. Urba S, Orringer MD, Turriss A, *et al.* Randomised trial of preoperative chemoradiation versus surgery alone in patients with locoregional esophageal carcinoma. *J Clin Oncol* 2001;**19:**305–13.

19. Bossett JF, Gigoux M, Triboulet JP, *et al.* Chemoradiotherapy followed by surgery compared with surgery alone in squamous-cell cancer of the oesophagus. *N Engl J Med* 1997;**337:**161–7.

20. Ono H, Kondo H, Gotoda T, *et al.* Endoscopic mucosal resection for treatment of early gastric cancer. *Gut* 2001;**48:**151–2.

21. Fujita H, Sueyoshi S, Yamana H*, et al.* Optimal treatment strategy for superficial esophageal cancer, mucosal resection versus radical esophagectomy. *World J Surg* 2001;**25:**424–31.

22. Tajima Y, Nakanishi Y, Ochiai A*, et al.* Histopathological findings predicting lymph node metastasis in patients with superficial esophageal carcinoma: analysis of resected tumours. *Cancer* 2000;**88:**1285–93.

23. Samuelsson L, Hambraeus GM, Mercke CE, Tylen U. CT staging of oesophageal carcinoma. *Acta Radiologica* 1984;**25:**7–11.

24. Sondenaa K, Skaane P, Nygaard K, Skjennald A. Value of computed tomography in preoperative evaluation of respectability and staging in oesophageal carcinoma. *Euro J Surg* 1992; **158:**537–40.

25. Van Overhagen H, Lmeris JS, Berger MY, *et al.* CT assessment of respectability prior to transhiatal oesophagectomy for oesophageal-gastroesophageal junctional carcinoma. *J Compt Assist Tomogr* 1993;**17:**367–73.

26. Trenker SW, Halvorsen RA, Thompson WM. Neoplasms of the upper gastrointestinal tract. *Radiol Clin N Am* 1994;**32:**15–24.

27. Botet JF, Lightdale CJ, Zauber AG, *et al.* Preoperative staging of esophageal cancer: comparison of endoscopic US and dynamic CT. *Radiology* 1991;**181:**419–25.

28. Rosch T. Endosonographic staging of esophageal cancer: a review of literature results. *Gastrointest Endosc Clin N Am* 1995;**5:**537–47.

29. Vickers J, Alderson D. Oesophageal cancer staging using endoscopic ultrasonography. *Br J Surg* 1998;**85:**994–8.

30. Richards DG, Brown TH, Manson JM. Endoscopic ultrasound in the staging of tumours of the oesophagus and gastro-oesophageal junction. *Ann R Coll Surg Engl* 2000;**82:**311–17.

31. Harris KM, Kelly S, Berry E*, et al.* Systematic review of endoscopic ultrasound in gastro-oesophageal cancer. *Health Technol Assess* 1998; 2.

32. Allum WH, Griffin SM, Watson A, Colin-Jones D. Guidelines for the management of oesophageal and gastric cancer. *Gut* 2002;**50**(Suppl. V):V1–V23.

33. Kelly S, Harris KM, Berry E, Hutton J, Roderick P, Cullingworth J*, et al.* A systematic review of the staging performance of endoscopic ultrasound in gastro-oesophageal carcinoma. *Gut* 2001;**49**:534–9.

34. Cella DF, Tulsky DS, Grey G, Sarafian B, Linn E, Bonomi A*, et al.* The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;**11**:570–9.

35. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use, 3rd edn. Oxford: Oxford University Press; 2003.

36. Scottish Audit of Gastro-Oesophageal Cancer. www.show.scot.nhs.uk/crag. Edinburgh: Scottish Executive Health Department; 2002.

Appendix 2 Hospital sites, principal investigators and trial practitioners

Gartnavel Hospital, Glasgow (later transferred to Glasgow Royal Infirmary): PI Mr Grant Fullarton; and trial practitioner Heather Hodgson.

Aberdeen Royal Infirmary: PI Professor Kenneth Park; and trial practitioner Dr Maureen Gillan.

Ninewells Hospital Dundee: PI Dr Elaine Henry; and trial practitioner Shirley Mcleod.

Gloucestershire Royal Infirmary: PI Professor Hugh Barr; and trial practitioners Clive Stokes, Alison Morgan and Gemma Ash.

Leicester Royal Infirmary: PI Dr Shona Campbell; and trial practitioners Nafisa Patel and Deborah Glancy.

Royal Blackburn Hospital: PI Dr Vishal Kaushik; and trial practitioners Lesley Miller and Vivienne Tickle.

Queen Alexandra's Hospital: Portsmouth, PI Dr Anthony Higginson; and trial practitioner Kay Bird.

Southampton General Hospital: PI Mr James Byrne; and trial practitioners Susan Horton and Julia Abab.

Appendix 3 Composition of oversight committees

Data Monitoring and Ethics Committee composition

Independent chairperson　　　　　Dr Hugh Gilmour, Consultant Pathologist, Royal Infirmary of Edinburgh.

Independent statistician　　　　　Professor Marion Campbell, Director, Health Services Research Unit, University of Aberdeen.

Patient representative　　　　　　Mr David Kirby, Oesophageal Patients Association, www.opa.org.uk.

Trial Steering Committee composition

Independent chairperson　　　　　Dr Robert Heading, Consultant Gastroenterologist, Royal Infirmary of Edinburgh.

Independent member　　　　　　　Professor Antoon Lerut, Consultant Surgeon, University Hospital Gasthuisberg, Leuven, Belgium.

COGNATE investigator　　　　　　Mr Grant Fullarton, Consultant Surgeon, Western Infirmary Glasgow.

COGNATE investigator　　　　　　Dr Shona Campbell, Consultant Radiologist, Leicester Royal Infirmary.

Chief investigator　　　　　　　　Professor Ian Russell, Swansea and Bangor Universities.

Chief investigator　　　　　　　　Professor Kenneth Park, Consultant Surgeon, Aberdeen Royal Infirmary.

Trial statistician　　　　　　　　Dr Daphne Russell, Swansea and Bangor Universities.

Trial manager　　　　　　　　　Ms Ceri Bray then Dr Angela Gliddon, NWORTH, Bangor University.

Co-opted investigator　　　　　　Mr Stephen Attwood, Consultant Surgeon, Northumbria Healthcare Foundation NHS Trust.

(Together with all members of the DMEC.)

Trial Management Team

Chief investigator　　　　　　　　Professor Ian Russell, Swansea and Bangor Universities.

Chief investigator　　　　　　　　Professor Kenneth Park, Consultant Surgeon, Aberdeen Royal Infirmary.

Trial statistician　　　　　　　　Dr Daphne Russell, Swansea and Bangor Universities.

Trial manager　　　　　　　　　Ms Ceri Bray then Dr Angela Gliddon, NWORTH, Bangor University.

Trial data manager　　　　　　　Ms Rhiannon Whitaker then Dr Angela Gliddon, NWORTH, Bangor University.

Outcome specialists   Dr David Ingledew, School of Psychology, and Ms Rhiannon Whitaker, NWORTH, Bangor University.

Health economists   Professor Rhiannon Tudor Edwards and Ms Seow Tien Yeo, Centre for Health Economics and Medicines Evaluation, Bangor University.

IT specialist   Mr Kevin Mawdsley, NWORTH, Bangor University.

Appendix 4 Further information on CONSORT diagram

Appendix 4.1 CONSORT eligibility: reasons for exclusion

Reasons for exclusion, numbered by order in which they appear in CONSORT diagram (*Figure 4*)		No. of people	%	Cumulative %
1)	Histology negative	28	2.4	2.4
2.1)	'Not fit for surgery'	375	32.6	35.0
2.2)	WHO status 3 or 4	8	0.7	35.7
3)	Metastases	174	15.1	50.8
4)	No consent (pre-MDM)	90	7.8	58.6
5)	No consent (post-MDM)	24	2.1	60.7
6.1)	In another trial	7	0.6	61.3
6.2.1)	EUS already done	32	2.8	64.1
6.2.2)	EUS already booked	4	0.3	64.4
6.3.1)	EUS needed (surgeon request)	18	1.6	66.0
6.3.2)	EUS needed (prior to surgery)	10	0.9	66.8
6.3.3)	EUS needed (specific reason)	18	1.6	68.4
6.3.4)	EUS needed (staging)	13	1.1	69.5
6.3.5)	EUS needed (management plan)	19	1.6	71.2
6.3.6)	EUS needed (med/MDM decision)	13	1.1	72.3
6.3.7)	EUS needed ('off trial')	29	2.5	74.8
6.4)	EUS not needed/EUS not suitable	8	0.7	75.5
6.5.1)	Organisational: geography	7	0.6	76.1
6.5.2)	Organisational: misunderstanding	9	0.8	76.9
6.5.3)	Organisational: time constraint	15	1.3	78.2
6.5.4)	Other organisational	15	1.3	79.5
6.5.5)	No longer eligible	8	0.7	80.2
6.5.6)	Not suitable (other)	3	0.3	80.5
7)	Randomised in error	2	0.2	80.6
8)	Randomised in trial	223	19.4	100.0
	TOTAL	1152	100.0	

Appendix 4.2 Eligibility by centre: all eight centres

Reasons for exclusion (see *Figure 4*)	Centre code: number (%)								
	20	21: Aberdeen	22	23: Gloucester	24	26	29	30	Total
Histology negative	2 (1)	0 (0)	2 (1)	0 (0)	1 (9)	3 (8)	15 (11)	5 (9)	28 (2)
'Not fit for surgery' or WHO 3 or 4	127 (44)	34 (20)	106 (37)	25 (15)	1 (9)	7 (18)	58 (44)	25 (46)	383 (33)
Metastases	76 (26)	30 (17)	42 (15)	2 (1)	1 (9)	13 (34)	6 (5)	4 (7)	174 (15)
No consent	15 (5)	23 (13)	14 (5)	20 (12)	2 (18)	1 (3)	25 (19)	14 (25)	114 (10)
In another trial	0 (0)	0 (0)	2 (1)	4 (2)	0 (0)	0 (0)	1 (1)	0 (0)	7 (1)
EUS already done or planned	3 (1)	4 (2)	20 (7)	1 (1)	0 (0)	0 (0)	8 (6)	0 (0)	36 (3)
EUS needed	39 (13)	1 (1)	60 (21)	0 (0)	0 (0)	10 (26)	10 (8)	0 (0)	120 (10)
EUS not needed or not suitable	3 (1)	2 (1)	2 (1)	0 (0)	0 (0)	0 (0)	1 (1)	0 (0)	8 (1)
Not suitable: other	11 (4)	8 (5)	29 (10)	1 (1)	0 (0)	2 (5)	6 (5)	0 (0)	57 (5)
Randomised[a]	**16** (5)	**72** (41)	**11** (4)	**110** (67)	**6** (55)	**2** (5)	**1** (1)	**7** (13)	**225** (20)
Total	**292** (100)	**174** (100)	**288** (100)	**163** (100)	**11** (100)	**38** (100)	**131** (100)	**55** (100)	**1152** (100)

a The two randomised in error were both from Aberdeen.

Appendix 4.3 Cumulative deaths and survivors among 223 patients

Time period	EUS deaths	Non-EUS deaths	Total deaths	Cumulative survivors censored at times up to…		
				EUS	Non-EUS	Total
Up to 1 month	1	2	3	1	1	2
Up to 3 months	3	7	10	2	1	3
Up to 6 months	15	17	32	2	3	5
Up to 12 months	33	36	69	2	3	6
Up to 18 months	42	53	95	16	12	28
Up to 24 months	51	61	112	23	21	44
Up to 36 months	60	71	131	33	26	59
Up to end of trial	61	74	135	50	38	88

Appendix 5 Further information on psychometric analyses

Appendix 5.1 Principal components analysis of FACT Specific Concerns at month 1 ($n = 173$)

Item	Component[a]						
	1	2	3	4	5	6	7
Item–component loadings							
Ga12. I have trouble swallowing food	0.86	0.18	−0.03	−0.01	0.20	0.05	0.01
[b]E1. I have difficulty swallowing solid foods	0.80	0.24	−0.06	−0.07	0.10	0.06	0.14
E5. I choke when I swallow	0.72	−0.04	0.04	0.11	0.03	0.13	−0.08
[b]E2. I have difficulty swallowing soft or mashed foods	0.69	0.17	−0.02	0.04	0.34	−0.18	−0.02
[b]HN7. I can swallow naturally and easily	−0.69	−0.12	−0.08	0.07	0.07	−0.02	0.09
HN1. I am able to eat the foods that I like	−0.64	−0.55	0.01	−0.07	0.05	−0.11	0.05
Ga6. I have discomfort or pain when I eat	0.55	0.27	0.23	0.03	0.42	0.02	0.12
E6. I am able to enjoy meals with family or friends	−0.52	−0.45	0.00	−0.18	0.13	−0.08	0.04
Ga1. I have a loss of appetite	0.13	0.79	0.12	0.08	0.12	0.10	−0.03
[b]C6. I have a good appetite	−0.26	−0.72	−0.19	0.04	0.01	0.05	0.22
[b]HN5. I am able to eat as much food as I want	−0.32	−0.65	−0.23	0.11	−0.01	−0.06	0.16
[b]Ga10. My digestive problems interfere with my usual activities	0.44	0.60	0.05	0.17	0.26	0.06	−0.09
C2. I am losing weight	0.12	0.57	0.03	−0.09	0.28	0.13	0.21
Ga4. I am bothered by a change in my eating habits	0.53	0.55	0.00	0.29	0.10	−0.03	−0.03
[b]Ga9. I avoid going out to eat because of my illness	0.45	0.55	−0.18	0.35	0.00	0.02	0.01
[b]Leu4. Because of my illness, I have difficulty planning for the future	0.12	0.47	0.18	0.43	0.09	0.04	0.06
HI12. I feel weak all over	0.06	0.45	0.26	0.36	0.07	0.33	−0.11
An2. I feel tired	−0.08	0.40	0.34	0.31	0.13	0.20	−0.03
[b]Ga5. I have a feeling of fullness or heaviness in my stomach area	−0.01	0.13	0.71	0.27	−0.05	0.17	−0.12
[b]ACT11. I have pain in my stomach area	−0.06	0.23	0.70	0.16	0.20	0.22	0.04
[b]C1. I have swelling or cramps in my stomach area	0.13	−0.15	0.64	0.33	0.07	−0.14	−0.14
HN2. My mouth is dry	0.01	0.38	0.56	−0.18	0.01	−0.06	−0.02
Ga14. I am bothered by gas (flatulence)	0.03	−0.05	0.20	0.70	−0.17	0.04	0.05
[b]Ga7. I worry about having stomach problems	0.04	0.29	0.13	0.60	0.36	0.09	0.01
E4. I have pain in my chest when I swallow	0.32	0.14	−0.05	−0.03	0.63	0.31	0.02
E3. I have difficulty swallowing liquids	0.50	0.11	0.17	−0.06	0.56	−0.18	−0.03

Item	Component[a]						
	1	2	3	4	5	6	7
Hep8. I have discomfort or pain in my stomach area	0.00	0.15	0.50	0.28	0.52	0.21	0.00
E7. I wake at night because of coughing	0.03	0.08	0.12	0.02	0.02	0.77	–0.11
Ga2. I am bothered by reflux or heartburn	0.01	0.11	0.02	0.31	0.27	0.54	–0.01
HN3. I have trouble breathing	0.18	0.06	0.48	–0.16	–0.09	0.50	0.08
HN10. I am able to communicate with others	–0.08	–0.05	–0.02	–0.04	0.03	–0.16	0.78
HN4. My voice has its usual quality and strength	0.00	–0.11	–0.01	0.13	0.05	0.02	0.65
C5. I have diarrhoea	–0.04	–0.04	0.25	0.34	0.32	–0.10	–0.49

a These seven components explain 61% of the variance.
b Items omitted from abridged combined scale.

Appendix 5.2 Principal components analysis of FACT Specific Concerns at month 3 (n = 150)

Item	Component[a]						
	1	2	3	4	5	6	7
Item–component loadings							
Ga12. I have trouble swallowing food	0.85	0.20	–0.03	0.13	0.07	–0.11	0.02
[b]E1. I have difficulty swallowing solid foods	0.76	0.27	–0.05	0.18	0.09	0.00	0.03
[b]E2. I have difficulty swallowing soft or mashed foods	0.73	0.32	0.01	0.06	0.07	0.13	–0.01
E5. I choke when I swallow	0.72	0.05	–0.01	0.13	–0.06	–0.28	0.08
E3. I have difficulty swallowing liquids	0.71	0.18	–0.09	0.29	0.04	0.08	–0.18
E4. I have pain in my chest when I swallow	0.66	0.06	0.35	–0.17	0.09	0.01	0.11
[b]HN7. I can swallow naturally and easily	–0.64	–0.33	0.15	–0.04	0.03	0.29	0.06
Ga4. I am bothered by a change in my eating habits	0.58	0.50	0.36	0.12	–0.04	0.02	–0.09
Ga6. I have discomfort or pain when I eat	0.53	0.28	0.52	0.10	0.01	0.04	0.07
[b]Ga10. My digestive problems interfere with my usual activities	0.47	0.46	0.43	0.30	–0.03	–0.04	0.09
[b]C6. I have a good appetite	–0.14	–0.83	–0.04	–0.13	–0.15	0.08	–0.03
E6. I am able to enjoy meals with family or friends	–0.24	–0.80	–0.06	–0.10	0.08	0.27	–0.06
Ga1. I have a loss of appetite	0.12	0.74	0.20	0.29	0.14	0.06	–0.06
[b]HN5. I am able to eat as much food as I want	–0.28	–0.71	–0.25	–0.16	–0.04	0.20	–0.01
[b]Ga9. I avoid going out to eat because of my illness	0.28	0.67	0.17	0.26	–0.13	0.00	0.06
HN1. I am able to eat the foods that I like	–0.49	–0.67	–0.07	–0.03	–0.06	0.15	–0.11
C2. I am losing weight	0.30	0.54	0.08	0.26	0.23	0.13	–0.16
Hep8. I have discomfort or pain in my stomach area	0.06	0.14	0.82	0.26	0.03	0.05	0.02
[b]ACT11. I have pain in my stomach area	0.13	0.06	0.77	0.16	0.14	–0.04	–0.04
[b]Ga5. I have a feeling of fullness or heaviness in my stomach area	–0.26	0.20	0.70	0.23	0.04	0.03	0.08

Item	Component[a]						
	1	2	3	4	5	6	7
[b]C1. I have swelling or cramps in my stomach area	–0.12	0.07	0.69	–0.01	0.20	–0.09	–0.03
[b]Ga7. I worry about having stomach problems	0.22	0.13	0.49	0.30	–0.28	–0.04	0.20
Ga14. I am bothered by gas (flatulence)	0.06	–0.04	0.45	0.40	–0.44	0.01	0.30
An2. I feel tired	0.15	0.21	0.21	0.75	0.07	–0.16	–0.03
HI12. I feel weak all over	0.12	0.29	0.35	0.70	0.04	–0.04	–0.09
HN2. My mouth is dry	0.11	0.20	0.20	0.62	0.12	0.29	–0.09
[b]Leu4. Because of my illness, I have difficulty planning for the future	0.19	0.28	0.13	0.59	0.04	–0.19	0.08
E7. I wake at night because of coughing	0.18	0.09	0.08	0.00	0.72	–0.06	0.09
HN3. I have trouble breathing	–0.03	0.05	0.24	0.24	0.70	–0.13	0.06
HN10. I am able to communicate with others	–0.09	–0.13	0.00	0.02	0.00	0.75	–0.12
HN4. My voice has its usual quality and strength	–0.07	–0.12	–0.09	–0.13	–0.23	0.68	0.19
Ga2. I am bothered by reflux or heartburn	0.05	0.17	0.35	0.06	0.25	–0.10	0.72
C5. I have diarrhoea	0.03	0.13	0.43	0.23	0.09	–0.19	–0.57

a These seven components explain 61% of the variance.
b Items omitted from abridged combined scale.

Appendix 5.3 Confirmatory factor analysis of FACT-G at month 1[a]

Item	Factor			
	Physical	Social	Emotional	Functional
Item–factor loadings				
GP1. I have a lack of energy	0.78			
GP2. I have nausea	0.52			
GP3. Because of my physical condition, I have trouble meeting the needs of my family	0.71			
GP4. I have pain	0.60			
GP5. I am bothered by side effects of treatment	0.57		0.21	
GP6. I feel ill	0.79			
GP7. I am forced to spend time in bed	0.73			
GS1. I feel close to my friends		0.87		
GS2. I get emotional support from my family		0.74		
GS3. I get support from my friends		0.91		
GS4. My family has accepted my illness		0.49		
GS5. I am satisfied with family communication about my illness		0.54		

Item	Factor			
	Physical	**Social**	**Emotional**	**Functional**
GS6. I feel close to my partner (or the person who is my main support)		0.38		
GE1. I feel sad			0.76	
GE2. I am satisfied with how I am coping with my illness			−0.46	
GE3. I am losing hope in the fight against my illness			0.42	
GE4. I feel nervous			0.79	
GE5. I worry about dying			0.59	
GE6. I worry that my condition will get worse			0.55	
GF1. I am able to work (include work at home)	−0.28			0.38
GF2. My work (include work at home) is fulfilling				0.62
GF3. I am able to enjoy life				0.86
GF4. I have accepted my illness		0.28		0.48
GF5. I am sleeping well				0.58
GF6. I am enjoying the things I usually do for fun				0.81
GF7. I am content with the quality of my life right now				0.83
Factor–factor correlations[b]				
Physical	–			
Social	−0.08	–		
Emotional	0.51	−0.31	–	
Functional	−0.65	0.31	−0.61	–

a $n = 173$. $\chi^2(282) = 453.93$, $p < 0.01$; SRMR = 0.07; CFI = 0.92; RMSEA = 0.06, p (close fit) = 0.06.

b Correlations between measurement errors: GP2 and GP3 negatively; GP4 and GP7 negatively; GS2 and GS5 positively; GS2 and GS6 positively; GS4 and GS5 positively; GE5 and GE6 positively; GF1 and GF2 positively; GF4 and GF6 negatively.

Appendix 5.4 Confirmatory factor analysis of FACT-G at month 3[a]

Item	Factor			
	Physical	**Social**	**Emotional**	**Functional**
Item–factor loadings				
GP1. I have a lack of energy	0.80			
GP2. I have nausea	0.58			
GP3. Because of my physical condition, I have trouble meeting the needs of my family	0.78			
GP4. I have pain	0.71			
GP5. I am bothered by side effects of treatment	0.53		0.29	
GP6. I feel ill	0.87			

Item	Factor			
	Physical	**Social**	**Emotional**	**Functional**
GP7. I am forced to spend time in bed	0.70			
GS1. I feel close to my friends		0.77		
GS2. I get emotional support from my family		0.64		
GS3. I get support from my friends		0.95		
GS4. My family has accepted my illness		0.63		
GS5. I am satisfied with family communication about my illness		0.52		
GS6. I feel close to my partner (or the person who is my main support)		0.54		
GE1. I feel sad			0.65	
GE2. I am satisfied with how I am coping with my illness			−0.16	0.32
GE3. I am losing hope in the fight against my illness			0.37	
GE4. I feel nervous			0.67	
GE5. I worry about dying	−0.26		0.89	
GE6. I worry that my condition will get worse			0.73	
GF1. I am able to work (include work at home)				0.70
GF2. My work (include work at home) is fulfilling				0.69
GF3. I am able to enjoy life		0.16		0.84
GF4. I have accepted my illness			−0.32	0.27
GF5. I am sleeping well				0.57
GF6. I am enjoying the things I usually do for fun				0.89
GF7. I am content with the quality of my life right now			−0.14	0.81
Factor–factor correlations[b]				
Physical	−			
Social	−0.17	−		
Emotional	0.45	−0.24	−	
Functional	−0.78	0.25	−0.41	−

a $n = 15$. $\chi^2(282) = 43.87$, $p < 0.01$; SRMR = 0.08; CFI = 93; RMSEA = 0.06, p (close fit) = 0.06.

b Correlations between measurement errors: GP1 and GP4 negatively, GP2 and GP3 negatively, GP3 and GP7 positively; GP4 and GP7 positively, GP6 and GP7 positively, GS2 and GS5 positively, GS3 and GS6 positively, GS4 and GS5 positively, GF1 and GF2 positively.

Appendix 5.5 Internal consistency and intercorrelations of scales at month 1 (*n* = 173)

Scale	Internal consistency	Correlations					
		1	2	3	4	5	6
1. FACT Specific Concerns	0.91	–					
2. FACT Physical	0.85	−0.52**	–				
3. FACT Social	0.85	−0.26**	0.16*	–			
4. FACT Emotional	0.78	−0.50**	0.41**	0.31**	–		
5. FACT Functional	0.86	−0.59**	0.63**	0.29**	0.48**	–	
6. EQ-5D	–	−0.43**	0.69**	0.11	0.39**	0.63**	–
7. EQ-VAS	–	−0.41**	0.61**	0.15*	0.41**	0.59**	0.55**

*$p < 0.05$, **$p < 0.01$.

Appendix 5.6 Internal consistency and intercorrelations of scales at month 3 (*n* = 150)

Scale	Internal consistency	Correlations					
		1	2	3	4	5	6
1. FACT Specific Concerns	0.93	–					
2. FACT Physical	0.89	−0.69**	–				
3. FACT Social	0.83	−0.18*	0.15	–			
4. FACT Emotional	0.75	−0.43**	0.41**	0.24**	–		
5. FACT Functional	0.90	−0.66**	0.72**	0.30**	0.41**	–	
6. EQ-5D	–	−0.51**	0.72**	0.17*	0.32**	0.68**	–
7. EQ-VAS	–	−0.63**	0.70**	0.17*	0.37**	0.66**	0.63**

*$p < 0.05$, **$p < 0.01$.

Appendix 5.7 FACT predicting EQ-5D at month 1
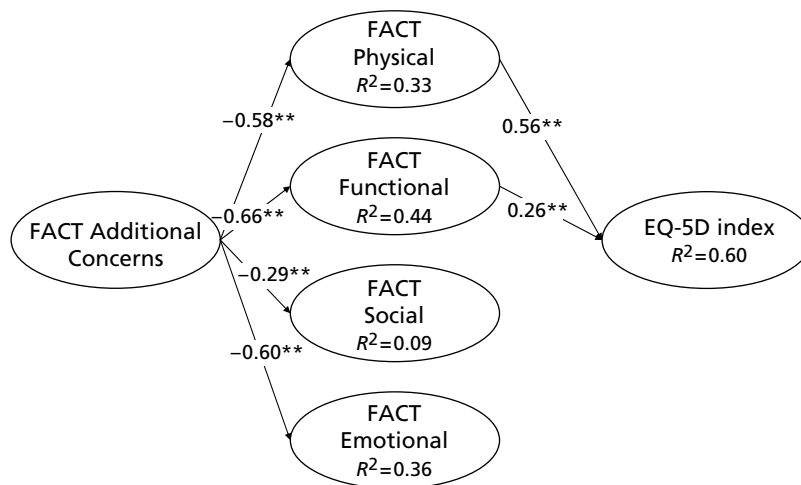
$\chi^2_3 = 3.90$; $p = 0.27$; SRMR = 0.02; CFI = 1.00; RMSEA = 0.04; p (close fit) = 0.45. **$p < 0.01$.

Appendix 5.8 FACT predicting EQ-5D at month 3
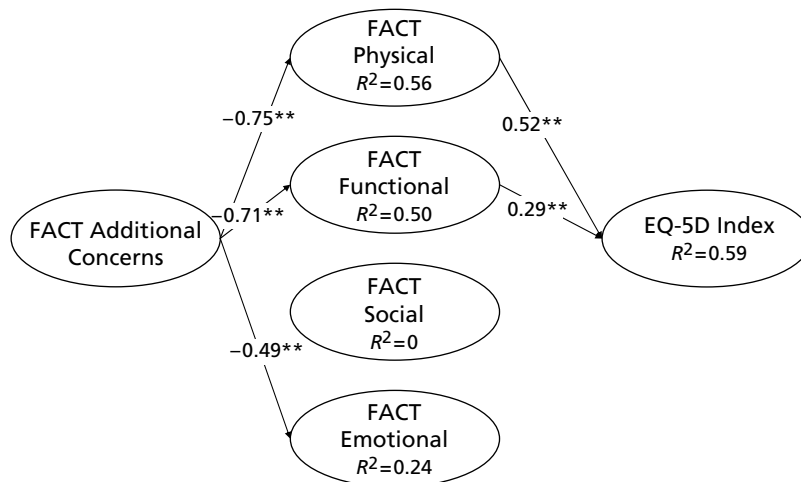
$\chi^2_4 = 9.50$; $p = 0.50$; SRMR = 0.06; CFI = 0.99; RMSEA = 0.10; p (close fit) = 0.14. **$p < 0.01$.

Appendix 5.9 FACT predicting EQ-VAS at month 1

FACT
Physical
$R^2=0.34$

FACT
Functional
$R^2=0.44$

FACT Additional
Concerns

FACT
Social
$R^2=0.09$

FACT
Emotional
$R^2=0.36$

VAS
$R^2=0.49$

−0.58** −0.66** −0.29** −0.60** 0.42** 0.33**
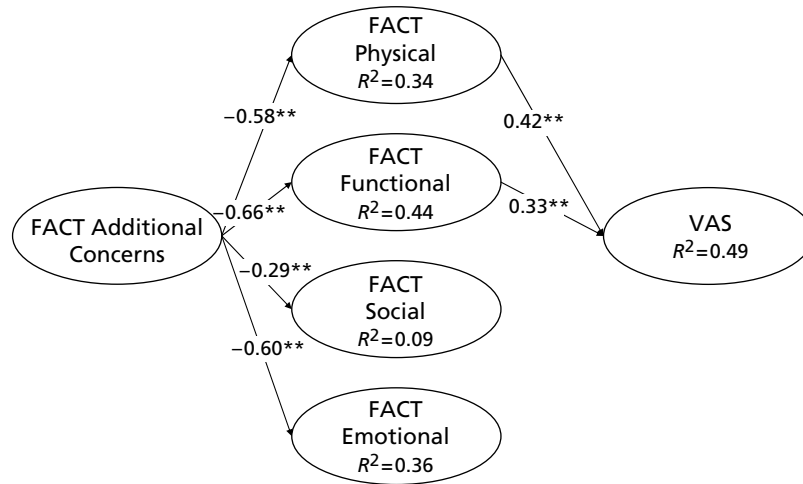
$\chi^2_4 = 6.52$; $p = 0.16$; SRMR = 0.04; CFI = 0.98; RMSEA = 0.07; p (close fit) = 0.21. **$p < 0.01$.

Appendix 5.10 FACT predicting EQ-VAS at month 3

FACT
Physical
$R^2=0.59$

FACT
Functional
$R^2=0.51$

FACT Additional
Concerns

FACT
Social
$R^2=0$

FACT
Emotional
$R^2=0.24$

VAS
$R^2=0.59$

−0.77** −0.72** −0.49** 0.54** 0.26**
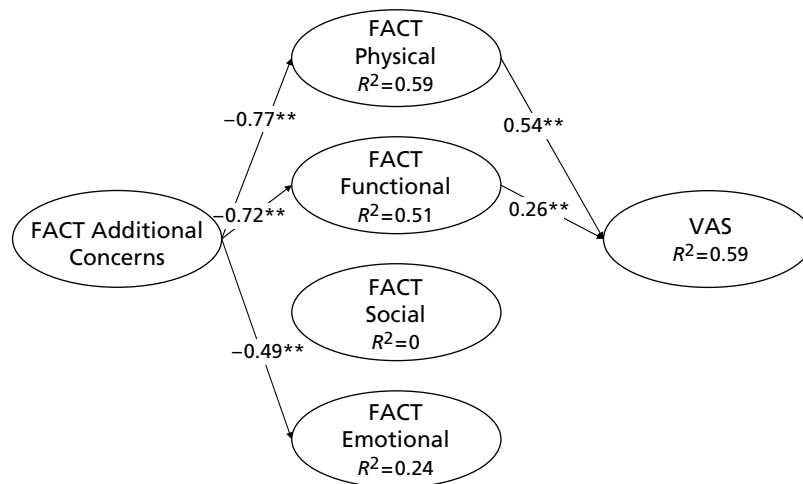
$\chi^2_4 = 6.35$; $p = 0.17$; SRMR = 0.06; CFI = 0.99; RMSEA = 0.06; p (close fit) = 0.33. **$p < 0.01$.

Appendix 6 Further information on clinical effectiveness

Appendix 6.1 Survival by allocated group: Cox regression comparing EUS and non-EUS groups (without covariates)

Model component or contrast	ß (coefficient in model)	Standard error of ß	Wald statistic	Significance[a]	Hazard ratio [exp(ß)]	95% CI for hazard ratio
EUS vs non-EUS	−0.277	0.173	2.563	0.109	0.758	0.540 to 1.064

a The significance level is slightly different from the significance of the log-rank test, which uses a different model. (Both are 'correct' but use different assumptions.)

Appendix 6.2 Survival by allocated group: Cox regression without interaction (including model coefficients)

Model component or contrast	ß (coefficient in model)	Standard error of ß	Wald statistic	df	Significance	Hazard ratio [exp(ß)]	95% CI for hazard ratio
EUS vs non-EUS	**−0.348**	**0.176**	**3.921**	**1**	**0.048**	**0.706**	**0.501 to 0.996**
Other centres vs (Aberdeen or Gloucester)	−0.073	0.128	0.328	1	0.567	0.929	0.723 to 1.194
Gloucester vs Aberdeen	−0.154	0.099	2.416	1	0.120	0.858	0.707 to 1.041
(T3, T4 or NR) vs (Tis, T1 or T2)	**0.303**	**0.107**	**7.962**	**1**	**0.005**	**1.353**	**1.097 to 1.670**
Multimodal plan vs not	**0.347**	**0.107**	**1.525**	**1**	**0.001**	**1.415**	**1.147 to 1.746**
Baseline EQ-5D	−0.335	0.484	0.477	1	0.490	0.716	0.277 to 1.849

Rows in bold type contribute significantly to the model.

Appendix 6.3 Baseline quality of life by centre for 213 participants in main analysis.

	Aberdeen (*n* = 70): mean (SD) minimum, maximum	Gloucester (*n* = 102): mean (SD) minimum, maximum	Other centres (*n* = 41): mean (SD) minimum, maximum
FACT (range 0–4, 4 best)			
Physical	3.44 (0.58) 0.6, 4.0	3.30 (0.59) 1.4, 4.0	3.48 (0.56) 1.6, 4.0
Social	3.47 (0.62) 1.0, 4.0	3.32 (0.70) 0.2, 4.0	3.60 (0.44) 2.0, 4.0
Emotional	3.05 (0.69) 0.7, 4.0	2.88 (0.82) 0.0, 4.0	3.03 (0.73) 1.2, 4.0
Functional	2.99 (0.84) 0.7, 4.0	2.72 (1.03) 0.2, 4.0	3.01 (0.94) 0.2, 4.0
General (average of four scores)	3.24 (0.42) 1.6, 4.0	3.06 (0.54) 1.5, 4.0	3.28 (0.49) 1.4, 4.0
FACT (range 0–4, 0 best)			
Additional Concerns (GOC)	1.12 (0.62) 0.2, 2.8	1.13 (0.66) 0.0, 2.6	1.01 (0.77) 0.0, 3.2
EQ-5D			
EQ-5D utility score (–0.6 to 1.0, 1.0 best)	0.83 (0.18) –0.1, 1.0	0.79 (0.18) 0.1, 1.0	0.81 (0.20) 0.3, 1.0
EQ-VAS (0–100, 100 best)	73.9 (16.1) 30, 100	68.8 (2.1) 20, 100	78.2 (16.6) 30, 100

GOC, gastro-oesophageal cancer.

Appendix 6.4 EQ-5D utility score by allocated group: unadjusted means at each time point (internally imputed; responders only)

	Non-EUS group		EUS group		Total
EQ-5D at	*n*	Mean (SD)	*n*	Mean (SD)	*n*
Baseline	105	0.800 (0.165)	104	0.810 (0.200)	209
1 month	82	0.768 (0.249)	90	0.757 (0.247)	172
3 months	74	0.675 (0.290)	75	0.676 (0.295)	149
6 months	66	0.658 (0.260)	70	0.660 (0.306)	136
12 months	**54**	**0.689 (0.309)**	**58**	**0.751 (0.195)**	**112**
18 months	37	0.735 (0.234)	40	0.690 (0.309)	77
24 months	23	0.657 (0.265)	26	0.695 (0.295)	49
36 months	12	0.588 (0.304)	13	0.651 (0.377)	45

Appendix 6.5　Twelve-month FACT scale scores and EQ-5D by allocated group: unadjusted means (internally imputed, responders only)

Scale	Non-EUS group (n = 54): mean (SD)	EUS group (n = 58): mean (SD)	EUS minus non-EUS: difference[a]
FACT			
Physical	3.10 (0.78)	3.13 (0.70)	0.036
Social	3.42 (0.70)	3.60 (0.60)	0.177
Emotional	3.21 (0.69)	3.28 (0.57)	0.065
Functional	2.51 (1.02)	2.52 (0.97)	0.012
General (average of four scores)	3.06 (0.61)	3.13 (0.57)	0.073
Additional Concerns	1.07 (0.68)	1.11 (0.73)	0.039
EQ-5D			
EQ-5D utility	0.689 (0.309)	0.751 (0.195)	0.062
EQ-VAS	72.8 (18.6)	73.1 (16.9)	0.30

a　There are no significant differences between groups.

Appendix 6.6　Twelve-month FACT scale scores, EQ-5D and EQ-VAS by allocated group: unadjusted means (imputed, survivors only: 70 non-endoscopic ultrasound and 73 endoscopic ultrasound group)

Scale	Non-EUS group: mean (SD)	EUS group: mean (SD)	EUS minus non-EUS: difference (95% CI)[a]
FACT			
Physical	3.08 (0.70)	3.08 (0.63)	0.01 (–0.22 to 0.23)
Social	3.40 (0.63)	3.56 (0.57)	0.16 (–0.04 to 0.36)
Emotional	3.23 (0.62)	3.23 (0.55)	0.01 (–0.19 to 0.20)
Functional	2.47 (0.93)	2.47 (0.93)	0.01 (–0.30 to 0.31)
General (average of four scores)	3.04 (0.55)	3.09 (0.56)	0.04 (–0.32 to 0.10)
Additional Concerns	1.09 (0.62)	1.16 (0.69)	0.08 (–0.14 to 0.29)
EQ-5D			
EQ-5D utility	0.68 (0.28)	0.74 (0.19)	0.06 (–0.02 to 0.14)
EQ-VAS	71.7 (16.8)	72.1 (16.2)	0.4 (–5.0 to 5.9)

a　There are no significant differences between groups.

Appendix 6.7 Quality of life by allocated group: unadjusted means, difference and confidence intervals from basic *t*-tests; each at each follow-up time point

Scale and time point	Non EUS group: mean (SD)	EUS group: mean (SD)	EUS minus non-EUS: difference (95% CI)		Significance
One month					
FACT-GP	2.98 (0.84)	2.94 (0.86)	−0.04	(−0.27 to 0.19)	0.724
FACT-GS	3.41 (0.77)	3.52 (0.69)	0.11	(−0.09 to 0.31)	0.268
FACT-GE	2.98 (0.80)	3.09 (0.78)	0.11	(−0.11 to 0.32)	0.330
FACT-GF	2.24 (0.96)	2.25 (1.07)	0.01	(−0.27 to 0.28)	0.962
FACT-G	2.90 (0.65)	2.95 (0.68)	0.05	(−0.14 to 0.23)	0.620
FACT-AC	1.29 (0.74)	1.16 (0.68)	−0.14	(−0.33 to 0.05)	0.159
EQ-5D	0.733 (0.285)	0.729 (0.265)	−0.005	(−0.079 to 0.070)	0.902
EQ-VAS	67.6 (21.3)	68.0 (18.2)	0.4	(−5.0 to 5.7)	0.885
3 months					
FACT-GP	2.73 (1.07)	2.78 (0.96)	0.04	(−0.23 to 0.32)	0.766
FACT-GS	3.21 (1.01)	3.44 (0.83)	0.23	(−0.012 to 0.48)	0.070
FACT-GE	2.83 (1.02)	3.01 (0.81)	0.17	(−0.08 to 042)	0.171
FACT-GF	1.98 (1.18)	2.05 (1.05)	0.07	(−0.24 to 0.37)	0.665
FACT-G	2.69 (0.93)	2.82 (0.75)	0.13	(−0.10 to 0.36)	0.271
FACT-AC	1.38 (0.94)	1.26 (0.73)	−0.12	(−0.35 to 0.10)	0.284
EQ-5D	0.615 (0.305)	0.658 (0.289)	0.043	(−0.037 to 0.123)	0.294
EQ-VAS	60.9 (26.6)	65.1 (20.8)	4.3	(−2.2 to 10.7)	0.193
6 months					
FACT-GP	2.31 (1.21)	2.31 (1.29)	0.00	(−0.34 to 0.34)	0.992
FACT-GS	2.89 (1.40)	3.06 (1.35)	0.17	(−0.20 to 0.54)	0.356
FACT-GE	2.55 (1.28)	2.65 (1.26)	0.10	(−0.24 to 0.44)	0.566
FACT-GF	1.70 (1.13)	1.84 (1.24)	0.14	(−0.18 to 0.46)	0.384
FACT-G	2.36 (1.14)	2.46 (1.18)	0.10	(−0.21 to 0.42)	0.512
FACT-AC	1.68 (1.04)	1.60 (1.05)	−0.08	(−0.36 to 0.21)	0.589
EQ5D	0.535 (0.323)	0.550 (0.347)	0.015	(−0.076 to 0.106)	0.744
EQ-VAS	55.5 (29.8)	56.7 (27.9)	1.2	(−6.6 to 9.0)	0.766
12 months					
FACT-GP	2.04 (1.56)	2.13 (1.53)	0.08	(−0.333 to 0.502)	0.691
FACT-GS	2.26 (1.69)	2.46 (1.71)	0.20	(−0.263 to 0.656)	0.401
FACT-GE	2.14 (1.61)	2.23 (1.56)	0.09	(−0.335 to 0.522)	0.667
FACT-GF	1.63 (1.39)	1.71 (1.38)	0.08	(−0.299 to 0.449)	0.692

Scale and time point	Non EUS group: mean (SD)	EUS group: mean (SD)	EUS minus non-EUS: difference (95% CI)		Significance
FACT0G	2.02 (1.51)	2.13 (1.50)	0.11	(−0.294 to 0.519)	0.587
FACT-AC	1.95 (1.31)	1.93 (1.30)	−0.02	(−0.373 to 0.331)	0.907
EQ-5D	0.449 (0.391)	0.509 (0.376)	0.061	(−0.043 to 0.164)	0.251
EQ-VAS	47.6 (36.5)	49.7 (36.0)	2.1	(−7.6 to 11.9)	0.666
18 months					
FACT-GP	1.63 (1.66)	1.68 (1.58)	0.05	(−0.43 to 0.53)	0.832
FACT-GS	1.71 (1.73)	1.81 (1.68)	0.10	(−0.40 to 0.61)	0.692
FACT-GE	1.77 (1.81)	2.00 (1.77)	0.23	(−0.30 to 0.76)	0.398
FACT-GF	1.36 (1.50)	1.40 (1.40)	0.04	(−0.39 to 0.47)	0.857
FACT-G	1.62 (1.62)	1.72 (1.56)	0.10	(−0.37 to 0.58)	0.660
FACT-AC	2.33 (1.36)	2.16 (1.34)	−0.15	(−0.55 to 0.25)	0.462
EQ-5D	0.377 (0.400)	0.394 (0.399)	0.017	(−0.102 to 0.135)	0.783
EQ-VAS	38.0 (39.2)	39.5 (37.1)	1.6	(−9.8 to 12.9)	0.785
24 months					
FACT-GP	1.16 (1.52)	1.33 (1.55)	0.16	(−0.33 to 0.66)	0.513
FACT-GS	1.29 (1.64)	1.52 (1.69)	0.23	(−0.31 to 0.77)	0.402
FACT-GE	1.34 (1.71)	1.62 (1.76)	0.29	(−0.28 to 0.85)	0.316
FACT-GF	0.93 (1.28)	1.13 (1.39)	0.20	(−0.24 to 0.63)	0.369
FACT-G	1.18 (1.50)	1.40 (1.55)	0.22	(−0.28 to 0.71)	0.382
FACT-AC	2.61 (1.28)	2.41 (1.38)	−0.21	(−0.68 to 0.27)	0.387
EQ-5D	0.251 (0.347)	0.330 (0.392)	0.079	(−0.041 to 0.198)	0.195
EQ-VAS	27.8 (36.0)	33.2 (37.8)	5.4	(−6.6 to 17.3)	0.376
36 months					
FACT-GP	1.02 (1.48)	0.96 (1.48)	−0.06	(−0.68 to 0.56)	0.847
FACT-GS	1.14 (1.64)	1.17 (1.62)	0.03	(−0.65 to 0.71)	0.929
FACT-GE	1.28 (1.80)	1.18 (1.66)	−0.10	(−0.82 to 0.62)	0.787
FACT-GF	0.76 (1.16)	0.81 (1.30)	0.05	(−0.46 to 0.57)	0.843
FACT-G	1.05 (1.50)	1.03 (1.47)	−0.02	(−0.64 to 0.60)	0.951
FACT-AC	2.78 (1.28)	2.75 (1.34)	−0.03	(−0.58 to 0.52)	0.923
EQ-5D	0.189 (0.312)	0.226 (0.373)	0.037	(−0.106 to 0.1)	0.607
EQ-VAS	24.6 (35.9)	26.2 (37.7)	1.6	(−13.8 to 16.9)	0.840

Appendix 6.8 Twelve-month FACT scale scores and EQ-5D by allocated group: unadjusted means (imputed, all cases)

Scale	Non-EUS group (*n* = 107): mean (SD)	EUS group (*n* = 106): mean (SD)	EUS minus non-EUS: difference (95% CI)
FACT			
Physical	2.04 (1.56)	2.13 (1.53)	0.08 (–0.33 to 0.50)
Social	2.26 (1.69)	2.46 (1.71)	0.20 (–0.26 to 0.66)
Emotional	2.14 (1.61)	2.23 (1.57)	0.09 (–0.33 to 0.52)
Functional	1.63 (1.39)	1.71 (1.38)	0.08 (–0.30 to 0.45)
General (average of four scores)	2.02 (1.51)	2.13 (1.50)	0.11 (–0.29 to 0.52)
Additional Concerns	1.95 (1.31)	1.93 (1.30)	–0.02 (–0.37 to 0.33)
EQ-5D			
EQ-5D utility	0.449 (0.391)	0.509 (0.376)	0.061 (–0.043 to 0.164)
EQ-VAS	47.6 (36.5)	49.7 (36.0)	2.1 (–7.6 to 11.9)

Appendix 6.9 Quality of life between allocated groups: full analysis of covariance comparisons, including effects of covariates but no interaction, at 12 months for EQ-5D

Model component or contrast	Type III sum of squares	df	Mean square	*F*	Significance
EUS vs non-EUS	0.176	1	0.176	1.258	0.263
Centre (three categories)	0.071	2	0.036	0.263	0.775
(T3, T4 or NR) vs (Tis, T1 or T2)	**0.991**	**1**	**0.991**	**7.095**	**0.008**
Age (years)[a]	0.502	1	0.502	3.596	0.059
Baseline EQ-5D	0.485	1	0.485	3.471	0.064
Error	28.777	206	0.140		
Corrected total	31.277	212			

a   A model with age replaced by multimodal management plan is nearly as good, but neither model is improved by using both age and multimodal plan

Rows in bold type contribute significantly to the model.

Appendix 6.10 Quality of life between allocated groups: full analysis of covariance comparisons, including effects of covariates but no interaction, at 12 months for FACT-G

Model component or contrast	Type III sum of squares	df	Mean square	F	Significance
EUS vs non-EUS	0.748	1	0.748	0.366	0.546
Centre (three categories)	2.754	2	1.377	0.673	0.511
(T3, T4 or NR) vs (Tis, T1 or T2)	**18.120**	**1**	**18.120**	**8.863**	**0.003**
Age (years)[a]	**9.768**	**1**	**9.768**	**4.778**	**0.030**
Baseline FACT-G	**24.941**	**1**	**24.941**	**12.199**	**0.001**
Error	421.170	207	2.045		
Corrected total	478.567	212			

a Even as a replacement for age, management plan does not contribute significantly to this model ($p = 0.12$)
Rows in bold type contribute significantly to the model.

Appendix 6.11 Quality of life between allocated groups: full analysis of covariance comparisons, including effects of covariates but no interaction, at 12 months for FACT-AC

Model component or contrast	Type III sum of squares	df	Mean square	F	Significance
EUS vs non-EUS	0.359	1	0.359	0.240	0.624
Centre (three categories)	0.774	2	0.387	0.260	0.772
(T3, T4 or NR) vs (Tis, T1 or T2)	**15.597**	**1**	**15.597**	**1.461**	**0.001**
Age (years)[a]	**6.454**	**1**	**6.454**	**4.329**	**0.039**
Baseline FACT-AC	**28.275**	**1**	**28.275**	**18.964**	**<0.001**
Error	307.141	206	1.491		
Corrected total	358.924	212			

a Even as a replacement for age, management plan does not contribute significantly to this model ($p = 0.25$)
Rows in bold type contribute significantly to the model.

Appendix 6.12 Quality of life between allocated groups: full analysis of covariance comparisons, including effects of covariates and interaction, at 12 months for EQ-5D

Model component or contrast	Type III sum of squares	df	Mean square	F	Significance
EUS vs non-EUS	0.178	1	0.178	1.325	0.251
Centre (three categories)	0.0144	2	0.072	0.536	0.586
(T3, T4 or NR) vs (Tis, T1 or T2)	**0.603**	**1**	**0.603**	**4.481**	**0.035**
Age (years)[a]	**0.582**	**1**	**0.582**	**3.974**	**0.039**
Baseline EQ-5D	**0.751**	**1**	**0.751**	**5.573**	**0.019**
Interaction (allocation group by baseline EQ-5D)	**1.206**	**1**	**1.206**	**8.966**	**0.003**
Error	27.571	205	0.135		
Corrected total	31.277	212			

a   Management plan does not significantly improve this model, although as a replacement for age it would be significant ($p = 0.048$)

Rows in bold type contribute significantly to the model.

Appendix 6.13 Quality of life between allocated groups: full analysis of covariance comparisons, including effects of covariates and interaction, at 12 months for FACT-G

Model component or contrast	Type III sum of squares	df	Mean square	F	Significance
EUS vs non-EUS	0.780	**1**	0.780	0.397	0.529
Centre (three categories)	3.306	2	1.653	0.842	0.432
(T3, T4 or NR) vs (Tis, T1 or T2)	**3.964**	**1**	**11.131**	**5.672**	**0.018**
Age (years)[a]	**11.362**	**1**	**11.362**	**5.790**	**0.017**
Baseline FACT-G	**20.499**	**1**	**20.499**	**10.446**	**0.001**
Baseline EQ-5D	0.450	1	0.450	0.229	0.633
Interaction (allocation group by baseline EQ-5D)	**20.834**	**1**	**20.834**	**10.617**	**0.001**
Error	400.333	204	1.962		
Corrected total	478.567	212			

a   Even as a replacement for age, management plan does not contribute significantly to this model ($p = 0.09$).

Rows in bold type contribute significantly to the model.

Appendix 6.14 Quality of life between allocated groups: full analysis of covariance comparisons, including effects of covariates and interaction, at 12 months for FACT-AC

Model component or contrast	Type III sum of squares	df	Mean square	F	Significance
EUS vs non-EUS	0.387	1	0.387	0.270	0.604
Centre (three categories)	1.473	2	0.736	0.513	0.599
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.120**	**1**	**1.120**	**7.051**	**0.009**
Age (years)[a]	**7.609**		**7.609**	**5.301**	**0.022**
Baseline FACT-AC	**28.148**	**1**	**28.148**	**19.611**	**< 0.001**
Baseline EQ-5D	0.309	1	0.309	0.215	0.643
Interaction (allocation group by baseline EQ-5D)	**14.322**	**1**	**14.322**	**9.978**	**0.002**
Error	292.807	204	1.435		
Corrected total	358.924	212			

a Even as a replacement for age, management plan does not contribute significantly to this model ($p = 0.23$).
Rows in bold type contribute significantly to the model.

Appendix 6.15 Survival adjusted by FACT-G, by allocated group: Cox regression

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS	0.724	0.514 to 1.020	0.065
Other centres vs (Aberdeen or Gloucester)	1.022	0.696 to 1.501	0.912
Gloucester vs Aberdeen	0.868	0.690 to 1.091	0.224
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.294**	**1.054 to 1.588**	**0.014**
Multimodal plan vs not	**1.409**	**1.142 to 1.739**	**0.001**
Baseline FACT-G	**0.617**	**0.428 to 0.888**	**0.009**

Rows in bold type contribute significantly to the model.

Appendix 6.16 Survival adjusted by FACT-G, by allocated group: Cox regression, including interaction

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS	0.729	0.517 to 1.029	0.073
Other centres vs (Aberdeen or Gloucester)	0.907	0.611 to 1.346	0.628
Gloucester vs Aberdeen	0.878	0.698 to 1.104	0.264
(T3, T4 or NR) vs (Tis, T1 or T2)	1.227	0.997 to 1.510	0.053
Multimodal plan vs not	**1.433**	**1.160 to 1.772**	**0.001**
Baseline FACT-G	**0.626**	**0.410 to 0.956**	**0.030**
Baseline EQ-5D	0.622	0.191 to 2.025	0.431
Interaction EQ-5D and group	**14.939**	**1.825 to 122.28**	**0.012**

Rows in bold type contribute significantly to the model.

Appendix 6.17 Survival adjusted by FACT-AC by allocated group: Cox regression

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS	0.769	0.545 to 1.084	0.134
Other centres vs (Aberdeen or Gloucester)	0.991	0.675 to 1.456	0.964
Gloucester vs Aberdeen	0.911	0.726 to 1.144	0.424
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.408**	**1.137 to 1.743**	**0.002**
Multimodal plan vs not	**1.422**	**1.153 to 1.754**	**0.001**
Baseline FACT-AC	**0.557**	**0.431 to 0.719**	**<0.001**

Rows in bold type contribute significantly to the model.

Appendix 6.18 Survival adjusted by FACT-AC by allocated group: Cox regression with interaction

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS	0.785	0.555 to 1.110	0.170
Other centres vs (Aberdeen or Gloucester)	0.868	0.584 to 1.289	0.482
Gloucester vs Aberdeen	0.938	0.745 to 1.180	0.584
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.324**	**1.066 to 1.645**	**0.011**
Multimodal plan vs not	**1.440**	**1.167 to 1.777**	**0.001**
Baseline FACT-AC	**0.534**	**0.407 to 0.700**	**<0.001**
Baseline EQ-5D	0.850	0.280 to 2.583	0.775
Interaction EQ-5D and group	**19.312**	**2.148 to 173.59**	**0.008**

Rows in bold type contribute significantly to the model.

Appendix 6.19 'Treatment received': Cox regressions for survival (*n* = 221)

Cox regression for survival: no covariates

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	0.753	0.534 to 1.062	0.106

Cox regression for survival: covariates, but no interactions

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	**0.697**	**0.492 to 0.986**	**0.041**
Other centres vs (Aberdeen or Gloucester)	1.013	0.690 to 1.488	0.946
Gloucester vs Aberdeen	0.843	0.671 to 1.058	0.141
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.361**	**1.102 to 1.680**	**0.004**
Multimodal plan vs not	**1.413**	**1.146 to 1.743**	**0.001**
Baseline EQ-5D	0.774	0.299 to 2.003	0.597

Cox regression for survival: with interaction between allocated group and baseline EQ-5D

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	**0.693**	**0.487 to 0.984**	**0.041**
Other centres vs (Aberdeen or Gloucester)	0.961	0.654 to 1.413	0.841
Gloucester vs Aberdeen	0.853	0.679 to 1.072	0.174
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.324**	**1.060 to 1.621**	**0.012**
Multimodal plan vs not	**1.440**	**1.167 to 1.778**	**0.001**
Baseline EQ-5D	0.685	0.255 to 1.841	0.453
Interaction EQ-5D and group	**7.403**	**0.972 to 56.393**	**0.053**

Appendix 6.20 'Treatment received': Cox regressions for survival adjusted by EQ-5D ($n = 213$)

Cox regression for survival adjusted by EQ-5D: no covariates

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	0.753	0.533 to 1.065	0.108

Cox regression for survival adjusted by EQ-5D: covariates, but no interactions

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	0.711	0.503 to 1.007	0.055
Other centres vs (Aberdeen or Gloucester)	1.039	0.708 to 1.526	0.844
Gloucester vs Aberdeen	0.879	0.699 to 1.104	0.267
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.327**	**1.073 to 1.641**	**0.009**
Multimodal plan vs not	**1.392**	**1.128 to 1.718**	**0.002**
Baseline EQ-5D	0.488	0.198 to 1.204	0.119

Cox regression for survival adjusted by EQ-5D: with interaction between allocated group and baseline EQ-5D

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	0.718	0.506 to 1.020	0.065
Other centres vs (Aberdeen or Gloucester)	0.964	0.674 to 1.457	0.964
Gloucester vs Aberdeen	0.887	0.706 to 1.115	0.305
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.280**	**1.032 to 1.587**	**0.024**
Multimodal plan vs not	**1.406**	**1.138 to 1.736**	**0.002**
Baseline EQ-5D	0.427	0.166 to 1.098	0.077
Interaction EQ-5D and Group	5.196	0.764 to 35.362	0.092

Appendix 6.21 'Treatment received': Cox regressions for survival adjusted by FACT-G (*n* = 213)

Cox regression for survival adjusted by FACT-G: no covariates

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	0.765	0.542 to 1.081	0.129

Cox regression for survival adjusted by FACT-G: covariates, but no interactions

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	**0.698**	**0.493 to 0.989**	**0.043**
Other centres vs (Aberdeen or Gloucester)	1.050	0.714 to 1.544	0.803
Gloucester vs Aberdeen	0.868	0.690 to 1.091	0.225
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.387**	**1.122 to 1.714**	**0.002**
Multimodal plan vs not	**1.405**	**1.139 to 1.733**	**0.002**
Baseline FACT-G	**0.609**	**0.422 to 879**	**0.008**

Cox regression for survival adjusted by FACT-G: with interaction between allocated group and baseline EQ-5D

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	0.704	0.495 to 1.001	0.050
Other centres vs (Aberdeen or Gloucester)	0.998	0.678 to 1.469	0.992
Gloucester vs Aberdeen	0.880	0.699 to 1.107	0.274
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.339**	**1.081 to 1.858**	**0.008**
Multimodal plan vs not	**1.423**	**1.152 to 1.757**	**0.001**
Baseline FACT-G	**0.609**	**0.396 to 0.938**	**0.024**
Baseline EQ-5D	0.898	0.278 to 2.901	0.857
Interaction EQ-5D and group	6.425	0.844 to 48.90	0.072

Appendix 6.22 'Treatment received': Cox regressions for survival adjusted by FACT-AC (*n* = 213)

Cox regression for survival adjusted by FACT-AC: no covariates

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	0.784	0.555 to 1.108	0.168

Cox regression for survival adjusted by FACT-AC: covariates, but no interactions

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	0.760	0.536 to 1.084	0.122
Other centres vs (Aberdeen or Gloucester)	0.991	0.674 to 1.456	0.961
Gloucester vs Aberdeen	0.917	0.730 to 1.144	0.454
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.415**	**1.142 to 1.743**	**0.001**
Multimodal plan vs not	**1.417**	**1.150 to 1.754**	**0.001**
Baseline FACT-AC	**0.556**	**0.431 to 0.719**	**<0.001**

Cox regression for survival adjusted by FACT-AC: with interaction between allocated group and baseline EQ-5D

Model component or contrast	Hazard ratio	95% CI	Significance
EUS vs non-EUS (treatment received)	0.766	0.539 to 1.089	0.138
Other centres vs (Aberdeen or Gloucester)	0.934	0.634 to 1.375	0.726
Gloucester vs Aberdeen	0.944	0.750 to 1.190	0.628
(T3, T4 or NR) vs (Tis, T1 or T2)	**1.363**	**1.098 to 1.692**	**0.005**
Multimodal plan vs not	**1.433**	**1.162 to 1.767**	**0.001**
Baseline FACT-AC	**0.533**	**0.405 to 702**	**<0.001**
Baseline EQ-5D	1.246	0.423 to 3.673	0.690
Interaction EQ-5D and group	7.414	0.919 to 59.84	0.060

Appendix 7 Further information on cost-effectiveness

Appendix 7.1 Mean frequency of contacts with secondary health care and hospital-prescribed drugs by 213 participants over 54 months

Time period (months)[a]	EUS group						Estimated mean frequency for complete case	Non-EUS group		
	Estimated mean frequency in complete interval							Estimated mean frequency in complete interval		
	0–12	12–18	18–24	24–36	36–48	48–60		0–12	12–18	18–24
Effective sample size n[b]	106.5	97.5	83	63	33.5	10		105	96	8.5
Outpatient visits										
Number of visits	11.31	2.47	2.90	2.97	1.28	0.80	**21.73**	11.29	2.06	1.45
Inpatient stay and day case										
Inpatient stay, cancer-related causes (no. of bed-days)	12.79	2.68	1.80	3.02	0.57	0.00	**2.84**	17.82	1.99	3.19
Inpatient stay, non-cancer-related causes (no. of bed-days)	0.49	0.24	0.00	0.03	0.00	0.00	**0.76**	0.18	0.03	0.04
Inpatient hospital stay, all causes (no. of bed-days)	13.28	2.91	1.80	3.05	0.57	0.00	**21.60**	18.00	2.02	3.23
Day-case EUS (no.)	0.90	0.00	0.00	0.00	0.00	0.00	**0.90**	0.03	0.00	0.00
Treatment										
Surgery (count, no. of bed-days)	0.61, 13.68	0.00	0.00	0.00	0.00	0.00	**0.61, 13.68**	0.62, 15.74	0.00	0.00
EMR day case	0.08	0.02	0.01	0.00	0.00	0.00	**0.11**	0.08	0.01	0.00
EMR inpatient (count, no. of bed-days)	0.03, 0.06	0.00	0.00	0.00	0.00	0.00	**0.03, 0.06**	0.01, 0.03	0.00	0.00
Chemotherapy (no. of cycles)	2.24	0.25	0.35	0.33	0.00	0.00	**3.17**	2.82	0.14	0.26
Radiotherapy (no. of fractions)	5.44	0.35	0.48	0.17	0.00	0.00	**6.44**	5.35	0.54	0.17

				Total sample						
			Estimated mean frequency for complete case	Estimated mean frequency in complete interval						Estimated mean frequency for complete case
24–36	36–48	48–60		0–12	12–18	18–24	24–36	36–48	48–60	
60	31	10		211.5	193.5	163.5	123	64.5	20	
1.75	1.16	0.00	**17.71**	11.30	2.27	2.19	2.37	1.22	0.40	**19.75**
3.38	3.16	0.00	**29.55**	15.29	2.34	2.48	3.20	1.81	0.00	**25.11**
0.08	0.00	0.00	**0.33**	0.34	0.13	0.02	0.06	0.00	0.00	**0.55**
3.47	3.16	0.00	**29.88**	15.62	2.47	2.50	3.25	1.81	0.00	**25.66**
0.00	0.00	0.00	**0.03**	0.47	0.00	0.00	0.00	0.00	0.00	**0.47**
0.00	0.00	0.00	**0.62, 15.74**	0.61, 14.70	0.00	0.00	0.00	0.00	0.00	**0.61, 14.70**
0.00	0.00	0.00	**0.09**	0.08	0.02	0.01	0.00	0.00	0.00	**0.10**
0.00	0.00	0.00	**0.01, 0.03**	0.02, 0.04	0.00	0.00	0.00	0.00	0.00	**0.02, 0.04**
0.13	0.13	0.00	**3.48**	2.53	0.19	0.31	0.24	0.06	0.00	**3.32**
0.43	0.00	0.00	**6.50**	5.39	0.44	0.33	0.30	0.00	0.00	**6.47**

Time period (months)[a]	EUS group						Estimated mean frequency for complete case	Non-EUS group		
	Estimated mean frequency in complete interval							Estimated mean frequency in complete interval		
	0–12	12–18	18–24	24–36	36–48	48–60		0–12	12–18	18–24
Effective sample size n[b]	106.5	97.5	83	63	33.5	10		105	96	8.5
Other procedures										
UGIE or dilatation day case (no.)	0.65	0.37	0.11	0.11	0.00	0.00	**1.24**	0.47	0.03	0.05
UGIE or dilatation inpatient (count, no. of bed-days)	0.09, 0.35	0.04, 0.11	0.01, 0.02	0.03, 0.08	0.00, 0.00	0.00, 0.00	**0.18, 0.56**	0.08, 0.33	0.01, 0.02	0.00, 0.00
STENT insertion day case	0.05	0.00	0.01	0.00	0.00	0.00	**0.06**	0.08	0.01	0.00
STENT insertion inpatient (count, no. of bed-days)	0.11, 0.47	0.02, 0.07	0.02, 0.11	0.05, 0.13	0.00, 0.00	0.00, 0.00	**0.20, 0.78**	0.06, 0.20	0.02, 0.07	0.02, 0.10
Other procedures day case	0.27	0.02	0.01	0.03	0.00	0.00	**0.34**	0.28	0.01	0.06
Other procedures inpatient (count, no. of bed-days)	0.10, 0.38	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00	**0.10, 0.38**	0.10, 0.64	0.00, 0.00	0.00, 0.00
Tests and investigations										
Count	66.03	8.69	7.41	7.63	2.90	4.60	**97.26**	78.00	7.34	7.48
PAMs										
Dietitian	4.86	0.41	0.01	0.08	0.15	0.00	**5.51**	6.25	0.02	0.40
Occupational therapist	0.08	0.02	0.04	0.00	0.00	0.00	**0.14**	0.05	0.02	0.00
Physiotherapist	5.65	0.09	0.01	0.08	0.00	0.00	**5.84**	9.01	0.06	0.16
Social worker	0.02	0.00	0.00	0.00	0.00	0.00	**0.02**	0.01	0.00	0.00
Other PAMs (e.g. district nurse, palliative care team)	1.03	0.04	0.07	0.40	0.00	0.00	**1.54**	2.13	0.22	0.29
All the above PAMs	11.65	0.56	0.13	0.56	0.15	0.00	**13.05**	17.45	0.32	0.84
Drugs										
Hospital prescribed drugs (no. of items)[c]	28.93	5.26	4.67	5.40	3.01	5.90	**53.18**	31.79	4.31	4.09

PAMS, Professionals Allied to Medicine; UGIE, upper gastrointestinal endoscopy.

a 0–12 months is the first year after randomisation; 12–18 months is the period between 12 and 18 months after randomisation, etc.

b The sample size decreases to include only those who were still in the trial: survivors observed throughout the interval, and those who died before the end of the interval but were randomised early enough for the end of the interval to occur before 31 July 2009. Participants (dead or alive) for whom the end of the trial occurred during the interval were included but given half of the weight of those in the trial for the full interval.

c Not including those prescribed for surgery or chemotherapy.

24–36	36–48	48–60	Estimated mean frequency for complete case	Total sample						Estimated mean frequency for complete case
				Estimated mean frequency in complete interval						
				0–12	12–18	18–24	24–36	36–48	48–60	
60	31	10		211.5	193.5	163.5	123	64.5	20	
0.22	0.00	0.00	**0.76**	0.56	0.20	0.08	0.16	0.00	0.00	**1.00**
0.00, 0.00	0.06, 0.16	0.00, 0.00	**0.15, 0.52**	0.09, 0.34	0.03, 0.07	0.01, 0.01	0.02, 0.04	0.03, 0.08	0.00, 0.00	**0.16, 0.54**
0.00	0.03	0.00	**0.12**	0.06	0.01	0.01	0.00	0.02	0.00	**0.09**
0.05, 0.45	0.03, 0.10	0.00, 0.00	**0.19, 0.92**	0.09, 0.34	0.02, 0.07	0.02, 0.10	0.05, 0.28	0.02, 0.05	0.00, 0.00	**0.19, 0.84**
0.05	0.06	0.00	**0.46**	0.27	0.02	0.04	0.04	0.03	0.00	**0.40**
0.00, 0.00	0.00, 0.00	0.00, 0.00	**0.10, 0.64**	0.10, 0.51	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00	**0.10, 0.51**
11.32	9.32	0.00	**113.46**	71.97	8.02	7.44	9.43	5.98	2.30	**105.15**
0.10	0.00	0.00	**6.77**	5.55	0.22	0.20	0.09	0.08	0.00	**6.14**
0.00	0.00	0.00	**0.07**	0.07	0.02	0.02	0.00	0.00	0.00	**0.11**
0.10	0.00	0.00	**9.33**	7.32	0.08	0.09	0.09	0.00	0.00	**7.57**
0.00	0.00	0.00	**0.01**	0.01	0.00	0.00	0.00	0.00	0.00	**0.01**
0.08	0.00	0.00	**2.72**	1.58	0.13	0.18	0.24	0.00	0.00	**2.13**
0.28	0.00	0.00	**18.90**	14.53	0.44	0.48	0.42	0.08	0.00	**15.96**
3.70	4.29	2.40	**5.58**	3.35	4.79	4.39	4.57	3.63	4.15	**51.87**

Appendix 7.2 Estimated mean cost of secondary health care and hospital prescribed drugs by 213 patients over 54 months

Time period (months)[a]	EUS group						Estimated mean cost (£) for complete case	Non-EUS group		
	Estimated mean cost (£) in complete interval							Estimated mean cost (£) in complete interval		
	0–12	12–18	18–24	24–36	36–48	48–60		0–12	12–18	18–24
Effective sample size n[b]	106.5	97.5	83	63	33.5	10		105	96	8.5
Total outpatient visits										
Number of visits	1290	262	324	318	133	70	**2397**	1317	226	165
Inpatient stay and day case										
Inpatient hospital stay (all cancer-related causes)	6003	973	741	1103	239	0	**9059**	8161	779	1366
Inpatient hospital stay (all non-cancer-related causes)	176	87	0	14	0	0	**278**	67	12	31
Inpatient hospital stay (all causes, cancer- and non-cancer-related)	6179	1060	741	117	239	0	**9336**	8228	791	1397
Day case EUS	497	0	0	0	0	0	**497**	16	0	0
Inpatient stay and day case costs										
Total cost	6676	1060	741	1117	239	0	**9833**	8244	791	1397
Treatment										
Surgery (all types including palliative surgery)	8226	0	0	0	0	0	**8226**	8976	0	0
EMR day case	39	10	6	0	0	0	**56**	40	5	0
EMR inpatient	25	0	0	0	0	0	**25**	13	0	0
Chemotherapy	1278	172	250	230	0	0	**1929**	1613	60	170
Radiotherapy	712	40	62	22	0	0	**836**	660	69	24
Other procedure										
UGIE or dilatation day case	339	187	55	54	0	0	**636**	245	16	25
UGIE or dilatation inpatient	154	48	10	33	0	0	**245**	148	9	0
Stent insertion day case	71	0	20	0	0	0	**91**	133	18	0
Stent insertion inpatient	346	55	75	107	0	0	**582**	158	56	72

				Total sample						
			Estimated mean cost (£) for complete case	Estimated mean cost (£) in complete interval						Estimated mean cost (£) for complete case
24–36	36–48	48–60		0–12	12–18	18–24	24–36	36–48	48–60	
60	31	10		211.5	193.5	163.5	123	64.5	20	
187	113	0	**2007**	1303	244	246	254	123	35	**2205**
1230	1556	0	**13,092**	7074	877	1049	1165	872	0	**11,037**
66	0	0	**176**	122	50	15	40	0	0	**227**
1296	1556	0	**13,268**	7196	927	1064	1205	872	0	**11,263**
0	0	0	**16**	258	0	0	0	0	0	**258**
1296	1556	0	**13,284**	7454	927	1064	1205	872	0	**11,521**
0	0	0	**8976**	8598	0	0	0	0	0	**8598**
0	0	0	**45**	40	8	3	0	0	0	**51**
0	0	0	**13**	19	0	0	0	0	0	**19**
82	76	0	**2001**	1445	116	211	158	36	0	**1965**
64	0	0	**816**	686	55	43	42	0	0	**826**
106	0	0	**392**	292	102	40	80	0	0	**514**
0	64	0	**221**	151	29	5	17	31	0	**233**
0	51	0	**201**	102	9	10	0	24	0	**145**
243	74	0	**603**	253	55	73	173	36	0	**590**

Time period (months)[a]	EUS group Estimated mean cost (£) in complete interval						Estimated mean cost (£) for complete case	Non-EUS group Estimated mean cost (£) in complete interval		
	0–12	12–18	18–24	24–36	36–48	48–60		0–12	12–18	18–24
Effective sample size n[b]	106.5	97.5	83	63	33.5	10		105	96	8.5
Other procedures day case	111	16	6	14	0	0	**147**	91	4	24
Other procedures inpatient	168	0	0	0	0	0	**168**	330	0	0
Total treatment cost	**11,470**	**528**	**484**	**460**	**0**	**0**	**12,942**	**12,405**	**237**	**315**
Total tests and investigations cost	**1003**	**103**	**124**	**97**	**28**	**44**	**1399**	**1041**	**109**	**126**
PAMs										
Dietitian	81	6	0.2	1	2	0	**91**	101	0.3	6
Occupational therapist	4	1	2	0	0	0	**6**	2	1	0
Physiotherapist	124	2	0.2	2	0	0	**128**	197	2	3
Social worker	0.7	0	0	0	0	0	**0.7**	0.4	0	0
Other PAMs (e.g. district nurse, palliative care team)	65	4	8	43	0	0	**120**	133	25	18
Total PAMs cost	**274**	**13**	**10**	**46**	**2**	**0**	**345**	**434**	**28**	**27**
Total NHS secondary sector costs	**20,714**	**1966**	**1683**	**2037**	**402**	**114**	**26,916**	**23,441**	**1392**	**2030**
Hospital drugs[c]										
Cost	997	178	162	379	269	227	**2212**	1816	201	165
Grand total cost (total secondary care cost + hospital prescribed drugs cost)	*21,710*	*2144*	*1844*	*2417*	*671*	*341*	*29,128*	*25,257*	*1592*	*2195*

PAMS, Professionals Allied to Medicine; UGIE, upper gastrointestinal endoscopy.

a 0–12 months is the first year after randomisation; 12–18 months is the period between 12 and 18 months after randomisation, etc.

b The sample size decreases to include only those who were still in the trial: survivors observed throughout the interval, and those who died before the end of the interval, but were randomised early enough for the end of the interval to occur before 31 July 2009. Participants (dead or alive) for whom the end of the trial occurred during the interval were included but given half of the weight of those in the trial for the full interval.

c Not including those prescribed for surgery or chemotherapy.

				Total sample						
			Estimated mean cost (£) for complete case	Estimated mean cost (£) in complete interval						Estimated mean cost (£) for complete case
24–36	36–48	48–60		0–12	12–18	18–24	24–36	36–48	48–60	
60	31	10		211.5	193.5	163.5	123	64.5	20	
19	26	0	**165**	101	10	15	17	13	0	155
0	0	0	**330**	248	0	0	0	0	0	248
514	291	0	**13,763**	11,934	384	401	486	140	0	13,345
132	77	0	**1485**	1022	106	125	114	52	22	1440
1	0	0	**109**	91	3	3	1	1	0	**100**
0	0	0	**3**	3	1	1	0	0	0	**4**
2	0	0	**205**	160	2	2	2	0	0	**166**
0	0	0	**0.4**	0.5	0	0	0	0	0	**0.5**
9	0	0	**184**	99	14	13	26	0	0	**152**
13	0	0	502	354	20	19	30	1	0	423
2141	2037	0	31,041	22,068	1681	1854	2088	1188	57	28,935
236	310	28	**2756**	1403	189	163	310	289	128	**2482**
2378	*2346*	*28*	*33,797*	*23,471*	*1870*	*2017*	*2398*	*1476*	*185*	*31,417*

Appendix 8 Further information on quality assurance

Appendix 8.1 Endoscopic ultrasound staging form

STAGING FORM – OESOPHAGEAL AND OESOPHAGOGASTRIC JUNCTIONAL CANCERS	
Cognate trial number:	**Scope(s) used**
Date	**Endoscopist**
Was video taken and stored? Y/N	**Supervisor**

1.   Endoscopic details:

Distance (cm) from incisor teeth to:

Proximal margin tumour	………………… cm	Hiatal hernia	Y/N
		from ………………… to ………………… cm	
Distal margin tumour	………………… cm	Stricture:	None or minimal
			Moderate or passable
			Tight or impassable
Tumour length	………………… cm	Dilatation?	Y/N
Location of OG junction	………………… cm		

2.   Other relevant data:

Prior anti-reflux surgery   Y/N   Prior gastric surgery   Y/N

3.   Tumour classification:

- Cervical (lower border cricoid to thoracic inlet)
- Intrathoracic   – upper (inlet to tracheal bifurcation)
                              – mid (bifurcation to just above OGJ)
- Lower thoracic/abdominal (including OGJ/intra-abdominal oesophagus)
               Type 1   Type 2   Type 3

4.   T staging: Position of aortic arch (cm)

Proximal margin tumour   ………………… cm   Proximal EUS margin   ………………… cm

Distal margin tumour   ………………… cm   Distal EUS margin   ………………… cm
                                                                                        from ………………… to ………………… cm

Longitudinal submucosal spread not visible at OGD? Y/N
               T1a   T1b   T2   T3   T4

Details of advanced T stage:
               'minimal' T3 (just breaches m. propria, 4th layer)   'bulky' T3 (extensive invasion beyond m. propria)

5.   N staging (please comment on all areas)**:**

Lymph node station	Total number seen	Total number suspicious	Total number involved

FNA performed   Y/N

6.   Staging summary (key overleaf):   **T ………………… N ………………… M…………………**

Signed ………………………………………… Status ………………………………………… Date …………………………………………

TNM Classification

Primary tumour (T)

TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma in situ
T1	Tumour invades lamina propria or submucosa
T2	Tumour invades muscularis propria
T3	Tumour invades adventitia
T4	Tumour invades adjacent structures

Regional lymph nodes (N)

NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastases
N1	Regional lymph node metastases

Distant metastases (M)

MX	Distant metastases cannot be assessed
M0	No distant metastases
M1	Distant metastases

Tumours of the lower thoracic oesophagus

M1a	Metastases in coeliac lymph nodes
M1b	Other distant metastases

Tumours in the mid thoracic oesophagus

M1a	Not applicable
M1b	Non-regional lymph nodes or other distant metastases

Tumours in the upper thoracic oesophagus

M1a	Metastases in cervical lymph nodes
M1b	Other distant metastases

Stage Classification

T	N	M	Stage
Tis	N0	M0	0
T1	N0	M0	I
T2	N0	M0	IIA
T3	N0	M0	IIA
T1	N1	M0	IIB
T2	N1	M0	IIB
T3	N1	M0	III
T4	Any N	M0	III
Any T	Any N	M1	IV
Any T	Any N	M1a	IVA
Any T	Any N	M1b	IVB

Appendix 8.2 Endoscopic ultrasound video assessment forms for internal assessors without access to EUS staging form

Video ID number ☐   Date of assessment ☐ / / ☐   Assessing centre ☐   Assessor ☐

Endoscopic results (pre-randomisation)

Distance (cm) from incisor teeth to:

Proximal margin tumour .... cm

Distal margin tumour .... cm

Tumour length .... cm

Location OG junction .... cm

Hiatal hernia?

Yes ☐

No ☐

From .... cm

To .... cm

Stricture?

None or minimal ☐

Moderate or passable ☐

Tight or impassable ☐

Adequacy of Video:

How good was the image quality?   Good ☐   Adequate ☐   Poor ☐

Was examination complete?   Complete ☐   Partial but adequate for staging ☐   Incomplete ☐

Comments: ....

Assessment:

Assessments	Your choice	How certain are you? (please tick)			Video times on which this choice is based	Comments
		Very	Fairly	Uncertain		
Tumour location						
T Classification						
N Classification						
M Classification						
Staging						

Comments: ....

Appendix 8.3　Endoscopic ultrasound video assessment form for external assessor with access to EUS staging form

Video ID number

Assessing centre

Date of assessment　 / /

Assessor

Adequacy of Video:

How good was the image quality?　Good ☐　Adequate ☐　Poor ☐

Was examination complete?　Complete ☐　Partial but adequate for staging ☐　Incomplete ☐

If partial or incomplete, why?

Comments ....

Adequacy of Assessment in Proforma:

Assessments	Choice in proforma	What do you think of this choice?			If disagree, your choice?	Time marks of video & Comments
		Agree	Reasonable	Disagree		
Tumour location						
T Classification						
N Classification						
M Classification						
Staging						

Was the assessment proforma fully completed?　Yes ☐　No ☐

Comments:

....

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library