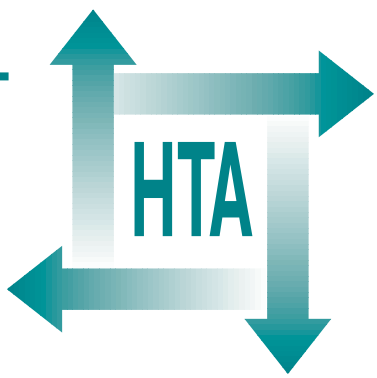
Health Technology Assessment 1998; Vol. 2: No. 18

Review

Systematic review of endoscopic ultrasound in gastro-oesophageal cancer

KM Harris
S Kelly
E Berry
J Hutton
P Roderick
J Cullingworth
L Gathercole
PJ O'Connor
JC Boyce
MA Smith

Health Technology Assessment
NHS R&D HTA Programme



Standing Group on Health Technology

Chair: Professor Sir Miles Irving,
Professor of Surgery, University of Manchester, Hope Hospital, Salford †

Dr Sheila Adam,
Department of Health

Professor Martin Buxton,
Professor of Economics, Brunel University †

Professor Angela Coulter,
Director, King's Fund, London

Professor Anthony Culyer,
Deputy Vice-Chancellor, University of York

Dr Peter Doyle,
Executive Director, Zeneca Ltd,
ACOST Committee on Medical Research & Health

Professor John Farndon,
Professor of Surgery, University of Bristol †

Professor Charles Florey,
Department of Epidemiology & Public Health, Ninewells Hospital & Medical School, University of Dundee †

Professor John Gabbay,
Director, Wessex Institute for Health Research & Development †

Professor Howard Glennester,
Professor of Social Science & Administration, London School of Economics & Political Science

Professor Sir John Grimley Evans,
Department of Geriatric Medicine, Radcliffe Infirmary, Oxford †

Dr Tony Hope,
The Medical School, University of Oxford †

Mr John H James,
Chief Executive, Kensington, Chelsea & Westminster Health Authority

Professor Richard Lilford,
Regional Director, R&D, West Midlands †

Professor Michael Maisey,
Professor of Radiological Sciences, UMDS, London

Dr Jeremy Metters,
Deputy Chief Medical Officer, Department of Health †

Mrs Gloria Oates,
Chief Executive, Oldham NHS Trust

Dr George Poste,
Chief Science & Technology Officer, SmithKline Beecham †

Professor Michael Rawlins,
Wolfson Unit of Clinical Pharmacology, University of Newcastle-upon-Tyne

Professor Martin Roland,
Professor of General Practice, University of Manchester

Mr Hugh Ross,
Chief Executive, The United Bristol Healthcare NHS Trust †

Professor Ian Russell,
Department of Health Sciences & Clinical Evaluation, University of York

Professor Trevor Sheldon,
Director, NHS Centre for Reviews & Dissemination, University of York †

Professor Mike Smith,
Director, The Research School of Medicine, University of Leeds †

Dr Charles Swan,
Consultant Gastroenterologist, North Staffordshire Royal Infirmary

Dr John Tripp,
Department of Child Health, Royal Devon & Exeter Healthcare NHS Trust †

Professor Tom Walley,
Department of Pharmacological Therapeutics, University of Liverpool †

Dr Julie Woodin,
Chief Executive, Nottingham Health Authority †

† Current members

HTA Commissioning Board

Chair: Professor Charles Florey, Department of Epidemiology & Public Health,
Ninewells Hospital & Medical School, University of Dundee †

Professor Ian Russell,
Department of Health Sciences & Clinical Evaluation, University of York *

Dr Doug Altman,
Director of ICRF/NHS Centre for Statistics in Medicine, Oxford †

Mr Peter Bower,
Independent Health Advisor, Newcastle-upon-Tyne †

Ms Christine Clark,
Honorary Research Pharmacist, Hope Hospital, Salford †

Professor David Cohen,
Professor of Health Economics, University of Glamorgan

Mr Barrie Dowdeswell,
Chief Executive, Royal Victoria Infirmary, Newcastle-upon-Tyne

Professor Martin Eccles,
Professor of Clinical Effectiveness, University of Newcastle-upon-Tyne †

Dr Mike Gill,
Director of Public Health and Health Policy, Brent & Harrow Health Authority †

Dr Jenny Hewison,
Senior Lecturer, Department of Psychology, University of Leeds †

Dr Michael Horlington,
Head of Corporate Licensing, Smith & Nephew Group Research Centre

Professor Sir Miles Irving
(Programme Director), Professor of Surgery, University of Manchester, Hope Hospital, Salford †

Professor Alison Kitson,
Director, Royal College of Nursing Institute †

Professor Martin Knapp,
Director, Personal Social Services Research Unit, London School of Economics & Political Science

Dr Donna Lamping,
Senior Lecturer, Department of Public Health, London School of Hygiene & Tropical Medicine †

Professor Theresa Marteau,
Director, Psychology & Genetics Research Group, UMDS, London

Professor Alan Maynard,
Professor of Economics, University of York †

Professor Sally McIntyre,
MRC Medical Sociology Unit, Glasgow

Professor Jon Nicholl,
Director, Medical Care Research Unit, University of Sheffield †

Professor Gillian Parker,
Nuffield Professor of Community Care, University of Leicester †

Dr Tim Peters,
Reader in Medical Statistics, Department of Social Medicine, University of Bristol †

Professor David Sackett,
Centre for Evidence Based Medicine, Oxford

Professor Martin Severs,
Professor in Elderly Health Care, Portsmouth University †

Dr David Spiegelhalter,
MRC Biostatistics Unit, Institute of Public Health, Cambridge

Dr Ala Szczepura,
Director, Centre for Health Services Studies, University of Warwick †

Professor Graham Watt,
Department of General Practice, Woodside Health Centre, Glasgow †

Professor David Williams,
Department of Clinical Engineering, University of Liverpool

Dr Mark Williams,
Public Health Physician, Bristol

Dr Jeremy Wyatt,
Senior Fellow, Health and Public Policy, School of Public Policy, University College, London †

* Previous Chair
† Current members

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Systematic review of endoscopic ultrasound in gastro-oesophageal cancer

KM Harris[1]
S Kelly[2]
E Berry[2]
J Hutton[3]
P Roderick[4]
J Cullingworth[1]
L Gathercole[1]
PJ O'Connor[1]
JC Boyce[2]
MA Smith[2]

[1] Department of Radiology, Leeds Teaching Hospitals NHS Trust, Leeds General Infirmary, Leeds, UK
[2] Institute of Medical Physics and Engineering and Centre of Medical Imaging Research, University of Leeds/Leeds Teaching Hospitals NHS Trust, Leeds General Infirmary, Leeds, UK
[3] MEDTAP International Inc., London, UK
[4] Wessex Institute of Health Research and Development, Southampton University, Southampton, UK

Published January 1999

This report should be referenced as follows:

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).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA web site (see overleaf).

NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Diagnostics and Imaging Panel and funded as project number 94/44/03.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for the recommendations for policy contained herein. In particular, policy options in the area of screening will, in England, be considered by the National Screening Committee. This Committee, chaired by the Chief Medical Officer, will take into account the views expressed here, further available evidence and other relevant considerations.

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.

Series Editors: Andrew Stevens, Ruairidh Milne and Ken Stein
Editorial Assistant: Melanie Corris

The editors have tried to ensure the accuracy of this report but cannot accept responsibility for any errors or omissions. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta

Contents

List of abbreviations

A	intercept of SROC (see section 3.1.5)
B	gradient of SROC (see section 3.1.5)
BIDS	Bath Information and Data Services
C	cardia†
CT	computed tomography
D	vertical axis of SROC
ERC	endoscopic retrograde cholangiography
ERCP	endoscopic retrograde cholangiopancreatography
EUS	endoscopic ultrasound
EWLS	equally weighted least squares (see section 3.1.5)
FN	false-negative
FNA	fine-needle aspiration
FP	false-positive
FPR	false-positive rate = 1 – specificity
FPR2	statistic resulting from the addition of 0.5 to each of TP, FN, FP and TN
ISI	Institute of Scientific Information
ISTAHC	International Society of Technology Assessment in Health Care
J	junction†
L	linear/curved†
M	miniprobe†
MeSH	medical subject heading
MRI	magnetic resonance imaging
N/A	not available†
NPV	negative predictive value
N/S	not stated†
O	oesophagus†
OCLC	Online Computer Library Centre
OR	odds ratio
OR2	statistic resulting from the addition of 0.5 to each of TP, FN, FP and TN
PPV	positive predictive value
Q*	Q star (see section 3.1.5)
QALY	quality-adjusted life years
QoL	quality of life
R	radial†
RCT	randomised controlled trial
ROC	receiver operator characteristic
RR	robust resistant (see section 3.1.5)
S	horizontal axis of SROC
Se	standard error
SIGLE	System for Information on Grey Literature
SROC	summary ROC curve
St	stomach†
TN	true-negative
TNM	tumour, node, metastasis
TP	true-positive
TPR	true-positive rate = sensitivity
TPR2	statistic resulting from the addition of 0.5 to each of TP, FN, FP and TN
UICC	International Union Against Cancer

† Used only in tables

Executive summary

Objectives

The aim was to review the literature relating to the use of endoscopic ultrasound for the preoperative staging of gastro-oesophageal cancer, especially regarding staging performance and staging impact. In addition, evidence was sought on the health economics, therapeutic impact and effect on patient outcome of endoscopic ultrasound in any clinical application.

Methods

Data sources

Electronic searches of MEDLINE and BIDS ISI formed the basis of the literature search. Other electronic resources searched included the Cochrane Library, EMBASE, Inside Information Plus, SIGLE and FirstSearch. Bibliographic listings of all retrieved articles were handsearched. Additionally, authors of abstracts, leading centres of endoscopic ultrasound, manufacturers and an endoscopic ultrasound e-mail discussion group were contacted with a request for unpublished information.

Study selection and validation

Study selection was a three-stage process using predefined inclusion and exclusion criteria. Only English language papers were included. The paucity of randomised controlled trials necessitated the acceptance of evidence from other study designs. For literature on staging performance, validation studies against a gold standard were included if there were sufficient numbers of patients and raw data were presented. For these studies, investigation of the validity of the evidence included analysis of the effect of the presence of any of 20 potential biases and the equipment and imaging protocol used.

Data extraction

Data were extracted from the studies selected using data extraction forms. Numerical values of staging performance for the completion of 2×2 contingency tables were extracted. Descriptive summaries were prepared for the other types of study where quantitative analysis was not feasible.

Data synthesis

Staging performance results (sensitivity, specificity, positive predictive value, negative predictive value, accuracy and odds ratio) were synthesised and receiver operator characteristic curves for the differentiation of tumour Stages T1 and T2 from T3 and T4 plotted. A summary statistic (Q*, balancing sensitivity and specificity) was read from the curve. Similar analysis for the discrimination of lymph node Stage N0 from N1 and above was performed.

Quantitative synthesis was not applicable for the studies of staging impact, therapeutic impact, patient outcome or health economics.

The robustness of the results was investigated by using regression techniques to incorporate bias risk and other factors (e.g. use of protocol) into the quantitative analysis.

Results

- Twenty-seven primary studies addressing the performance of endoscopic ultrasound for the preoperative staging of gastro-oesophageal cancer satisfied the inclusion criteria.
- The performance of endoscopic ultrasound in T staging gastro-oesophageal cancer was Q* = 0.91. For gastric T staging Q* = 0.93 and for oesophageal T staging Q* = 0.89.
  - The value for Q* was significantly ($p < 0.05$) lower for studies performed in the 1990s than for those in the 1980s.
  - The presence of stenosis resulting in non-traversability was found slightly, but significantly ($p < 0.05$), to reduce the staging performance of endoscopic ultrasound.
  - Radial probes performed better than linear probes in staging gastric cancer, although, in staging oesophageal cancer, there was no significant difference in the performance between probes.
- The performance of endoscopic ultrasound in N (lymph node) staging associated with gastro-oesophageal cancer was Q* = 0.79. For N staging associated with gastric cancer this was Q* = 0.76 and for N staging associated with oesophageal cancer Q* = 0.82.

  - Studies that reported attempts to perform some form of blinding achieved a significantly ($p < 0.05$) better performance compared with those that did not.
- Insufficient information for data synthesis was found on M staging (staging of metastases) and grouped TNM staging.
- There was insufficient information on the use of miniprobes (for subanalysing T1 tumours).
- There was little information about the use of fine-needle aspiration specifically applicable to gastro-oesophageal cancer.
- Eight studies compared the staging performance of endoscopic ultrasound with that of incremental computed tomography (CT), but the CT aspects of these were poorly performed and no measure of the staging impact of endoscopic ultrasound (EUS) could be determined.
- There was very little evidence regarding therapeutic impact, patient outcome and health economics.

Conclusions

- EUS is highly effective for the discrimination of Stages T1 and T2 from T3 and T4, in both the oesophagus and the stomach.
- Initial indications are that the performance for T staging at the cardia is less good.
- Non-traversable stenosis does reduce the staging performance of EUS, but evidence on whether this reduction justifies the risk of dilatation was not available.
- The studies available on the use of miniprobes report a high performance for discrimination between mucosal and submucosal cancer. No evidence regarding the subsequent impact of these findings is available.
- Lymph node staging with EUS has a lower performance than that of tumour staging.
- Staging for metastases using EUS alone is not satisfactory.

Recommendations

The following research recommendations were made by the authors:

- methodological research into the effect of searching only the major electronic databases and into factors that make publication bias less likely
- continued collaboration between reviewers in fields lacking randomised controlled trials regarding the assessment of study quality
- updating of this review, especially with regard to the proportion of non-traversable tumours encountered
- a study to determine the value of miniprobes prior to endoscopic mucosal resection
- well-designed studies, using the optimal protocols for both EUS and CT, to compare staging performance, which must also investigate the complementary use of the modalities
- further investigation of the use of fine-needle aspiration in gastro-oesophageal cancer in a study concentrating on lymph nodes
- retrospective studies to confirm the limited learning curve data currently available
- new studies, specifically designed to measure staging impact, therapeutic impact and patient outcome, because evidence in these areas is not currently available
- use of decision-modelling techniques to combine outcome and cost data from the new studies and other sources
- encouragement of imaging scientists both to perform better designed studies and to ensure that descriptions published in the literature are comprehensive.

Chapter 1

Background information

The references in this review are indicated in the text by superscripted numerals. They are divided into three sections:

- references cited in the review: superscripts with **no brackets**
- references included in the review: superscripts with **square brackets**
- references excluded from the review: superscripts with **round brackets.**

Although it is not the usual style for *Health Technology Assessment,* most of the **subheadings** have been decimalised. This is primarily to allow easy referral (especially from chapters 3, 4, 5, 8 and 9) to the five levels of the Fineberg evaluative framework that are discussed later in this chapter and which are central to the layout of this review.

In this opening chapter, the technology of endoscopic ultrasonography (EUS) is introduced and an overview of its potential strengths and weaknesses discussed. The tumour, node, metastasis (TNM) staging classification system is explained. This is followed by a description of the Fineberg framework for the evaluation of imaging technologies. This framework has been used as the basis for classifying studies in the rest of the review.

Endoscopic ultrasonography

EUS combines endoscopy and high-frequency ultrasound to allow a unique opportunity to visualise the gastrointestinal wall and adjacent structures. It has been in use since the early 1980s but has been slow to gain acceptance in certain countries, including the UK.

Equipment

There are two basic types of echoendoscope commercially available with either radial or linear array transducer technology. Linear array scanners provide a slightly more limited field of view and anatomical orientation is said by many to be more difficult, but they do however enable real-time ultrasound image-guided fine-needle biopsy to be undertaken. These instruments also have colour Doppler® facility for assessment of blood flow and the easier identification of blood vessels. EUS image-guided fine-needle biopsy has been described with conventional radial scanning echoendoscopes, but this should become easier with the introduction of a dedicated radial 'puncture scope'. Miniprobes are small, higher-frequency probes (e.g. 20 MHz) that can be passed down the biopsy channel of a conventional endoscope. Their high frequency ensures excellent resolution but the depth of penetration is restricted. This, combined with their more limited durability and relative cost, has contributed to their failure to gain more widespread acceptance. Most miniprobes utilise mechanical radial scanning technology, although linear miniprobes have been used. The radial scanning echoendoscopes currently commercially available use dedicated EUS processing equipment as a 'stand-alone' unit although they can share a common light source and optical processing unit with other conventional endoscopes. Most miniprobes require similar dedicated equipment. Linear or curved array echoendoscopes, although requiring an endoscopic light source, may be compatible with certain existing ultrasound processing units. This enables the ultrasound unit to be used at other times for conventional transabdominal scanning. Upper gastrointestinal endosonography is performed in a similar fashion to conventional endoscopy. The examination is usually performed using sedation as a day-case/out-patient episode.

Applications

EUS has been used in a wide range of clinical settings, including the staging of gastrointestinal tumours. The normal bowel wall appears on EUS as a five-layered structure *(Figure 1.1)*. Higher-frequency probes are able to identify up to nine distinct layers, including the circular and longitudinal muscle components of the muscularis propria. Echoendoscopes are available for use in the evaluation of the upper gastrointestinal tract and also for colorectal imaging. Adjacent organs and vascular structures are easily identified and act as useful landmarks; they are also important in the evaluation of the possible spread of malignancy. Upper gastrointestinal EUS allows visualisation of the pancreas, biliary tree, and the left and part of the right lobe of the liver. EUS has therefore not surprisingly been used in the evaluation of suspected pancreaticobiliary disease. Although there are certainly less potential procedure-related

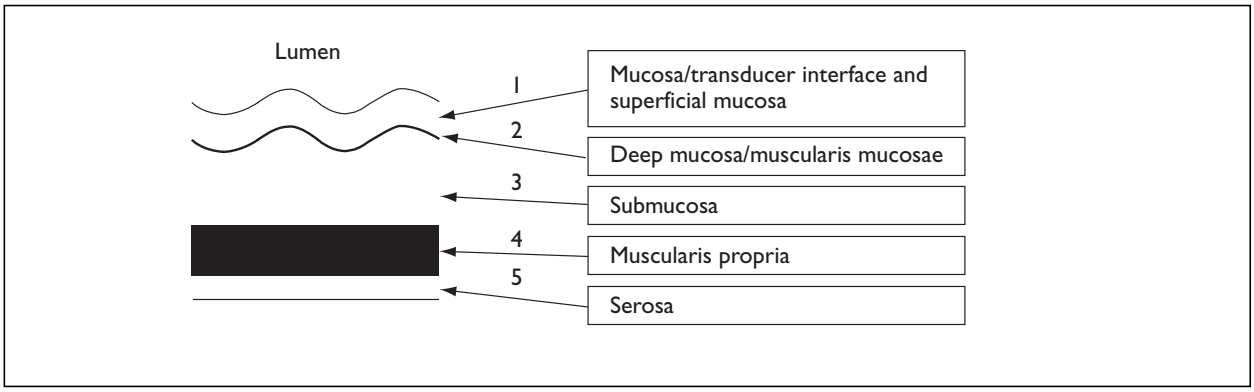


FIGURE 1.1 *Diagrammatic representation of the normal five-layered pattern of the bowel wall*

complications (e.g. acute pancreatitis) with EUS than endoscopic retrograde cholangiopancreatography (ERCP), its role has yet to be defined in relation to evolving less invasive imaging modalities, including magnetic resonance imaging (MRI). Most users would agree that EUS continues to provide useful information in the assessment of patients with gastro-oesophageal malignancy and published reports recommend its continued use. With this in mind, this review aims critically to review the available data in this area to define the role of EUS.

Staging of gastro-oesophageal cancer

In 1995 the death rate per million population in England and Wales from all malignant neoplasms was 2795 for men and 2498 for women. Of these, 137 and 84 were from oesophageal tumours in men and women respectively, and 165 and 104 were from gastric tumours, representing approximately 10% of all cancer deaths (dominated by lung and breast). Significantly, the death rate from oesophageal cancer has risen by 40% in men and 20% in women since 1980, while the death rate from gastric cancer is falling.[1]

Carcinomas of the oesophagus are either adenocarcinomas or squamous cell carcinomas; those of the stomach are adenocarcinomas. The area that surrounds the gastro-oesophageal junction is known as the cardia or the gastro-oesophageal junction *(Figure 1.2)*. In the UK, in common with many other countries, the natural history of gastro-oesophageal cancer appears to be changing in that a greater number of tumours are presenting at the cardia. There seems to be a change in the most frequent anatomical location of adenocarcinomas from the oesophagus and stomach to the cardia. It is often unclear if the tumour has arisen from metaplastic epithelium in the lower third of the oesophagus or from gastric mucosa. This trend has important implications

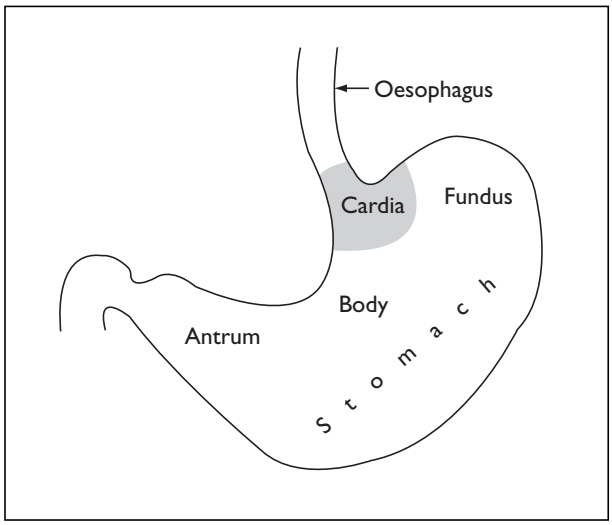


FIGURE 1.2 *The anatomical regions of the oesophagus and stomach*

because, clearly, any imaging modality must be well suited to evaluation of this area. The anatomy of the gastro-oesophageal junction can cause problems with EUS owing to oblique scanning through the bowel wall as it turns to become gastric fundus and body. Similar problems are encountered with axial computed tomography (CT) imaging and clearly the relative impact on the staging of tumours in this region needs full evaluation. The ability to identify the component layers of the bowel wall provides the basis for tumour staging within the widely accepted TNM classification. The International Union against Cancer (UICC) TNM classification[2] defines the extent of malignant tumours and allows easy correlation of results from more than one centre. The definition has recently (1997) been changed,[3] but all the studies included in this review use the 1987 definition described here. Although broadly similar, there are important differences between the TNM staging for oesophageal carcinoma and gastric carcinoma, as outlined below:

Oesophageal carcinoma:
T1: tumour invades lamina propria or submucosa
T2: tumour invades muscularis propria
T3: tumour invades adventitia
T4: tumour invades adjacent structures
N0: no regional lymph node metastasis
N1: regional lymph node metastasis
M0: no distant metastasis
M1: distant metastasis (including distant lymph node metastasis)

Gastric carcinoma:
T1: tumour invades lamina propria or submucosa
T2: tumour invades muscularis propria or subserosa (a tumour may penetrate the muscularis propria with extension into the gastrocolic or gastrohepatic ligaments or the greater omenta without perforation of the visceral peritoneum covering these structures and remain T2*; if the visceral peritoneum is breached it becomes T3)
T3: tumour penetrates serosa (visceral peritoneum) without invasion of adjacent structures
T4: tumour invades adjacent structures
N0: no regional lymph node metastasis
N1: metastasis in perigastric lymph node(s) within 3 cm of the edge of the primary tumour
N2: metastasis in perigastric lymph node(s) more than 3 cm from the edge of the primary tumour or in lymph nodes along the left gastric, common hepatic, splenic or coeliac arteries
M0: no distant metastasis
M1: distant metastasis

Direct invasion into an adjacent organ (T4) can be difficult to predict by using CT or EUS. Invasion is suspected when tumour is contiguous with an adjacent structure and it is considered likely when the tumour involves a significant percentage of its surface. Similar degrees of contiguity are employed during EUS assessment but, in addition, use can be made of the real-time nature of EUS: a tumour may be in close contact with an adjacent organ but clear movement between the two during respiration excludes invasion. This is particularly obvious with gastric cancers abutting the liver.

Local tumour invasion (T stage) is illustrated diagrammatically in *Figure 1.3*. Accurate staging of gastro-oesophageal tumours is essential to allow a well-informed decision to be made to plan appropriate treatment. Improvements in non-surgical management of advanced gastrointestinal tumours demand accurate staging. Such precise stage-dependent management would hopefully limit the incidence of unnecessary exploratory surgical interventions. Accurate tumour staging is also clearly important when comparing outcomes of various non-surgical interventions, as there is no pathological gold standard. At the other end of the disease spectrum there is also a requirement for accurate local tumour staging. Small superficial early gastrointestinal cancers can sometimes be removed endoscopically, but knowledge of the precise depth of tumour penetration and exclusion of more distant spread are essential prerequisites. EUS has a limited depth of penetration and, although it is well suited to the evaluation of local invasion, it is of limited usefulness in the overall assessment of more distant spread. Other imaging modalities such as CT, MRI and transabdominal ultrasound will without question be better suited to the evaluation of the possibility of distant tumour spread. CT, for example, can evaluate for distant spread the whole chest, abdomen and pelvis of a patient with malignancy, but may be less accurate in certain areas such as evaluation of local tumour and nodal status. Certainly, even state-of-the-art CT scanners are unable to resolve the component layers of the bowel wall and therefore cannot discriminate between T1 and T2 tumours. It is therefore also possible that subtle penetration of the oesophageal muscularis propria may not be

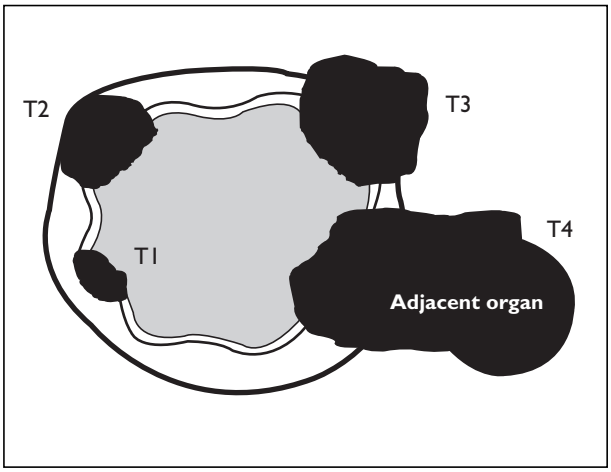


FIGURE 1.3 *Gastro-oesophageal T staging as demonstrated by EUS*

* The omental reflections around the stomach are not clearly seen with EUS and this classification raises important issues for EUS staging of gastric carcinomas. It is difficult or impossible to know if a carcinoma has penetrated the muscularis propria into the greater or lesser omenta but not breached the visceral peritoneum beyond (i.e. ?T2 or ?T3).

visible and therefore the accuracy of CT in discriminating T3 from T2 or less invasive tumours must be questioned. Unfortunately, in the UK, patients with gastro-oesophageal tumours often present late in the course of the disease and evaluation of the presence of distant metastases is an essential part of the staging process. It seems unlikely therefore that EUS will completely replace other imaging techniques in a significant proportion of patients. However, with improvements in chemoradiotherapeutic regimens (either before, after or as an alternative to surgical resection), accurate staging is essential to guide management. The accuracy of EUS in the staging of gastro-oesophageal cancer and the relationship with other imaging modalities, in particular CT, needs defining.

Tumour stenosis and non-traversable tumours

The most widely used echoendoscope is approximately 13 mm in diameter and larger than most diagnostic gastroscopes. Therefore, although a full evaluation of the oesophagus and stomach may have been possible with the conventional endoscope at the time of diagnosis, the subsequent EUS examination may be limited by an inability to negotiate a tight oesophageal or, less commonly, gastric stenosis. Even with non-traversable tumours, some information will be obtained from the proximal extent of the tumour and adjacent lymph nodes but clearly this may only be the 'tip of the iceberg'. A smaller, 8.5 mm diameter 'oesophagoprobe' without viewing optics is available and will negotiate a greater number of oesophageal tumours. This single-frequency (7.5 MHz) radial scanning echoendoscope is introduced over a guide wire previously manipulated through the lumen of the tumour. Miniprobes will negotiate all but the tightest of strictures but their poor depth of ultrasound penetration impairs imaging. There is a suggestion that the configuration of the linear ultrasound transducer enables greater penetration into the non-traversable tumour than the radial probe, which scans perpendicular to the long axis of the endoscope, but it is unclear if this adds significant further information. Image interpretation may be affected by compression of the tumour when it is just traversable with some resistance. Such compression may efface the component layers of the oesophageal wall, with the potential risk of tumour over-staging. Dilatation of malignant oesophageal strictures may allow subsequent passage of an echoendoscope. Such dilatation could be performed in advance of or at the time of EUS. Dilatation of oesophageal tumours is not without risk; oesophageal perforation is a potentially serious complication. The benefits of a complete EUS examination versus the increased morbidity and mortality from the dilatation of usually advanced tumours must be considered.

Fine-needle aspiration

Fine-needle aspiration (FNA) for cytology is a technique that is used for the diagnosis of a wide range of pathologies, including breast cancer. With superficial tumours such as this the needle can be introduced into the palpable abnormality either 'freehand' or utilising imaging such as ultrasound to target the lesion. The presence of malignant cells in the aspirate confirms the diagnosis but, because of the possibility of sampling error from such a small biopsy, a negative result must be viewed with extreme caution. Although the diagnosis of gastro-oesophageal cancer may be suspected clinically or on barium contrast studies, histological confirmation is required. The mucosal abnormality is usually easily recognised during upper gastrointestinal endoscopy, and conventional biopsies suffice to establish the histopathological diagnosis. EUS-guided FNA would therefore not appear to have a role in the initial diagnosis of gastro-oesophageal cancer but may contribute to the overall staging process. CT and MRI use lymph node size as the major determinant of possible malignant involvement but lymph nodes can become significantly enlarged in a number of benign conditions. The prevalence of enlarged benign lymph nodes varies considerably, depending in part upon coexisting disease, such as granulomatous disorders, which may also have a worldwide geographical distribution, and previous occupational exposure such as pneumoconiosis. EUS also relies upon lymph node size to predict malignant involvement but can evaluate other echo characteristics in addition. Four features have been shown consistently to correlate well with malignant involvement as assessed by EUS:

- size: diameter greater than 1 cm
- rounded
- well-defined margin
- homogeneous hypoechoic echotexture.

When all features are present, the node is highly likely to be involved but with three or fewer criteria fulfilled there is uncertainty. The confirmation of malignant lymph node involvement should, in certain situations, have important management implications. If a coeliac axis lymph node is shown to be involved by tumour spread from an upper oesophageal cancer, the TNM stage becomes M1. However, involved nodes immediately adjacent to the tumour may be removed at the time of surgical resection and their presence will not directly

influence management. Clearly only nodes adjacent to the bowel or within easy reach can be sampled. EUS-guided FNA has been used in an attempt to confirm malignancy in suspicious nodes using a purpose-built needle (usually 25 to 21 gauge). A positive (malignant) result confirms involvement but there is the possibility of false reassurance from a negative (no malignant cells) sample. EUS FNA has been used in an attempt to establish the diagnosis of various other submucosal masses but here again the main problem is one of sample size. Pathologists often require a larger biopsy to establish the true nature of certain lesions such as stromal tumours. The possible increase in procedure-related morbidity and mortality from the addition of FNA to an EUS examination needs defining.

EUS equipment is costly and requires considerable experience in both endoscopic techniques and ultrasound image interpretation. It is an endoscopic technique requiring sedation and, as with conventional gastroscopy, has recognised procedure and sedation-related complications, morbidity and mortality. Some patients, even with sedation, will not be able to tolerate the procedure and there is therefore clearly a need to define the important relationships with other less invasive imaging modalities that are currently available. These techniques, in particular CT, can stage gastro-oesophageal tumours but it is important to determine if EUS can cost-effectively add further useful information to improve patient outcome.

The Fineberg evaluative framework

In the evaluation of imaging technologies it is customary to use the framework first proposed by Fineberg and colleagues[4] in a study of CT scanning. This identifies the conceptual stages that must be considered before concluding that the use of a different diagnostic or staging method has influenced patient outcome. The structure of this report is based on the modified[5] Fineberg levels, which are:

- technical capability (not covered in this review)
- diagnostic or staging performance (or accuracy)
- diagnostic or staging impact
- therapeutic impact
- patient outcome
- health economics.

Evidence at each stage is needed because success at one level does not guarantee impact at the next. Improved diagnostic accuracy may not change diagnosis but only increase the confidence of clinicians in the correctness of their diagnosis. A change in diagnosis may not alter the therapeutic approach. For example, a change of staging of a cancer from III to IV may not alter therapy, but a change from Stage III to Stage II could lead to much more active intervention. Changes in therapy may not affect the patient's ultimate outcome in terms of length of survival. However, the use of more patient-centred outcome measures, such as impact on quality of life (QoL), is more likely to lead to some perceived change in outcome as therapy is altered.

Economists have added a further stage to the Fineberg framework, covering investigations related to the costs associated with the technique. A new diagnostic method may lead to improvements in patient outcomes but the extra cost of its use may not be justified, given the alternative uses to which the necessary healthcare resources could be put.

Because study designs for each level of the framework differ, the criteria used for assessing the validity of primary studies vary from level to level of the framework. The actual criteria used, and our reasons for selecting them, are detailed in chapter 3. In the following paragraphs an overview of the important features of study design at each level is presented.

1. Staging performance

To assess staging accuracy, the comparison should be made against a gold standard reference test, which should be applied to all patients. Study designs at this level need to be rigorous and free from bias to ensure validity. At this level, the analogy to a well-perfomed randomised controlled trial (RCT) is a comparison to a gold standard test, with patients randomly allocated to the study and with blinding between tests. In medical imaging many studies do not adhere to these standards. Many imaging studies address the clinical effectiveness of the technology and in this scenario, where the performance in clinical routine is being evaluated, the appropriate study design is not as well defined. There is a vast amount of literature on study design; an article by the authors of this review has been published, which describes potential biases prevalent in performance studies of medical imaging modalities.[6] A further issue is the question of reproducibility because the utility of a test that does not give the same result on repeat examinations, or for different observers, is doubtful. Intraobserver reproducibility is a

function of experience and may be assessed by comparing the staging accuracy of individuals as they begin to use the technology and again when they have achieved competence.

2. Staging impact

The focus of the analysis here is whether the use of the new diagnostic technology leads to any patient receiving a different diagnosis. Diagnostic impact can also be defined in terms of the confidence of the clinicians in their diagnosis. A more confident diagnosis can have two effects: active therapy may be undertaken more quickly, and fewer confirmatory, duplicated diagnostic tests may be used, reducing the cost of the diagnostic process.

To assess diagnostic impact, studies must determine how the results of tests are used by clinicians in reaching a diagnosis and how they fit into a sequence of clinical decisions. This requires a more pragmatic or naturalistic design than that of an experimental study addressing diagnostic performance. Studies of diagnostic impact can be designed so that a study group receives the test under investigation and a control group does not. Randomisation is still desirable to prevent selection bias, but blinding the clinician to the source of information is usually not possible nor is it necessarily desirable. In situations where tests are complementary rather than directly substitutive, the problem may be to determine the optimal sequence of testing. This can be done by randomising patients between two predetermined access routes to the first test and allowing clinicians to request the second test if desired. The latter may better reflect how the tests may be subsequently used in routine practice but does introduce the possibility of selection bias for the second test.

In the absence of studies designed intentionally to evaluate diagnostic impact, those that use comparative technologies in a diagnostic performance study design can be used as a secondary standard. This will provide an experimental comparison between the performance statistics of the competing tests, but it does not supply any information regarding the subsequent impact of replacing the existing technology. In addition, the possibility of the tests being complementary may be overlooked.

3. Therapeutic impact

This is defined in terms of changes in the clinical management of patients as a result of diagnosis by a different modality. This can involve changes between curative and palliative therapy, surgical or medical management, or a faster introduction of the same therapy. As for diagnostic impact, the basic factors of good study design are important, including initial randomisation. Sufficiently extended follow-up to observe changes in management is preferable. Studies often record intentions to treat only, particularly if the patient has received two tests and clinicians are asked to assess the impact of each one independently.

In order to evaluate the therapeutic impact of a technology, such rigorous methodologies are required because simply recording the end measure can be misleading. For example, if EUS was found to improve differentiation between a resectable and an unresectable tumour when compared with that of CT, the test in current clinical use, then this influence will be identified by more patients being assigned to the therapeutic protocol with the best prognosis (i.e. surgery) and fewer to other protocols such as radiotherapy, chemotherapy or palliative therapy. However, if only the number of patients proceeding to surgery versus the other protocols was recorded for both CT and EUS, any observed differences could be purely due to the patient characteristics of those receiving each test (i.e. an increased number proceeding to surgery for EUS could be due to those patients having early stage cancers, of which more are resectable). Therefore, studies reporting only these end outcomes were identified but they were not considered to evaluate therapeutic impact and hence were not assessed.

4. Patient outcome

The culmination of diagnostic and therapeutic impact is a change in patient outcome. The follow-up period required to verify this will vary with the disease area. The outcome measure chosen must be appropriate to the question being addressed and the analytical approach used. If the risk of morbidity and mortality associated with the tests under study differ, then the outcome measure must be able to incorporate these effects. The timing and frequency of outcome assessment must also take this into account. Again, studies not following strict methodologies to evaluate patient outcome, but supplying limited information regarding patient survival, were identified but were not assessed.

Many studies do not record systematically the impact of a new staging or diagnostic approach on actual diagnoses or on subsequent therapeutic decisions. This makes assessment of the potential impact on outcomes very difficult, regardless of the outcome measure chosen. Many clinical studies of therapeutic intervention stop short of measuring actual outcomes. For example, the success of cancer therapy is often judged on the tumour

response rate, without long-term follow-up to see whether patients with a good response to treatment actually survive longer. Survival is not the only aspect of outcome of interest to patients. A large literature has built up on the assessment of QoL during and after treatment.[7]

Some researchers have developed guidelines for judging the relevance and validity of QoL studies.[8] These are helpful but relate only to one aspect of outcome measurement. There is a hierarchy of patient outcomes, as described by Fries and Singh[9] and discussed in some of the economics guidelines (e.g. Drummond and Jefferson[10]). Economists working on the evaluation of health have developed measures of outcome that combine the change in both quantity (survival) and QoL, the most frequently used being the quality-adjusted life-year (QALY).[11] There is still much disagreement and debate about the best way to assess the impact on QoL for such outcome measures. For example, should instruments be disease specific or generic? Should respondents be patients with the condition or representative samples of the general population? The following could be regarded as a widely accepted ranking of outcome measures, moving from the least useful to the most broadly applicable:

- intermediate clinical outcomes (e.g. tumour response rates or number of true-positive diagnoses)
- final clinical outcomes (survival rates)
- cumulative clinical outcomes (life years saved)
- patient-assessed outcomes (QoL)
  - patient satisfaction
  - disease-specific QoL scales
  - generic QoL scales
- patient preference measures (combining QoL and survival)
  - QALYs
  - healthy year equivalents
- monetary values of patient benefits
  - willingness to pay.

The final category of monetary measures would allow economists to carry out a full cost–benefit analysis of health care, expressing costs and benefits in monetary terms. In healthcare systems where patients do not buy care directly, willingness to pay has to be elicited by indirect means such as conjoint analysis. These approaches have been widely used in other areas of economics, such as transport and the environment, and are increasingly the subject of new research in health economics. Their application in the imaging field has been limited to assessments of acceptability of different tests to patients (e.g. willingness to pay for more expensive tests that involve less risk or discomfort) but their wider potential has been recognised.[12]

5. Health economics

The focus of economic evaluation is on resource use and benefits to patients that may be realised in a routine healthcare delivery situation. In other words, the external validity of studies is more important than the internal validity, which means that RCTs have drawbacks as vehicles for economic evaluation. Nevertheless, the criteria for judging economic studies can still be grouped into four main categories: study design; data collection; analysis; and interpretation of results.

Conclusion

In this chapter, the background to the clinical issues involved in the use of EUS for the staging of gastro-oesophageal cancer has been described. Additionally, the Fineberg evaluative framework has been outlined. This framework will be used in the review to provide a logical structure for classifying studies. In the next chapter, the questions to be addressed within the review are clarified.

Chapter 2

Hypotheses tested in the review

EUS in gastro-oesophageal cancer

Should EUS staging be recommended for gastro-oesophageal tumours?

Conventional EUS probes

- What is the staging performance of EUS for distinguishing Stages T1 and T2 from Stages T3 and T4 for tumours of the oesophagus, at the cardia, of the stomach, or of the three sites overall?
- What is the staging performance of EUS for distinguishing Stage N0 from Stage N1 for lymph nodes associated with primary tumours of the oesophagus?
- What is the staging performance of EUS for distinguishing Stage N0 from Stages N1 and above for lymph nodes associated with primary tumours at the cardia, of the stomach, or of the three sites overall?
- Is there any evidence regarding the staging performance of EUS for M staging for metastases from primary tumours of the oesophagus, at the cardia, of the stomach, or of the three sites overall?
- Is there sufficient evidence to assess the grouped TNM staging performance of EUS for tumours of the oesophagus, at the cardia, of the stomach, or of the three sites overall?
- What proportion of patients (in the staging performance studies) are incomplete for EUS owing to tumour stenosis and how does this influence the result?
- What information is available regarding reproducibility and the learning curve?

EUS miniprobes

- What is the staging performance of EUS miniprobes for distinguishing mucosal from submucosal T1 tumours of the oesophagus, at the cardia, of the stomach, or of the three sites overall?

Is there any evidence that EUS has an impact on methods used for staging?

- Is there any evidence directly comparing CT/MRI/positron emission tomography and EUS for staging?
- Is there any evidence directly comparing EUS miniprobes with conventional EUS?
- Is there any evidence about the value of EUS-guided FNA in gastro-oesophageal cancer?
- Is there any evidence directly comparing dedicated EUS machines with EUS probes designed to be used with existing ultrasound equipment?

EUS in any clinical application

What evidence is available about EUS at the therapeutic impact, patient outcome and health economics levels of the Fineberg framework?

- Is there any evidence about the therapeutic impact of EUS?
- Is there any evidence about the effect on patient outcome of the use of EUS?
- Is there any evidence on the health economics of EUS?

Chapter 3
Review methods

A multidisciplinary review team was assembled; its composition was designed to:

- ensure a broad spread of relevant expertise
- minimise the potential for bias in the review
- facilitate dissemination of both review methodology and review results among several professional groups.

The panel comprised the authors and an external member, who is an opinion leader in the field of radiology. The professions represented were medical physics, radiology, radiography, health economics and public health medicine. The representative from radiology is a dedicated gastrointestinal radiologist who is in constant touch with both gastroenterologists and gastrointestinal surgeons.

The review methodology was broadly based on that recommended in Centre for Reviews and Dissemination Report 4.[13] The approach to validity assessment was adapted for the different levels of the evaluative framework introduced by Fineberg,[4] and outlined in chapter 1. For the levels of staging performance and staging impact, the particular clinical applications in gastro-oesophageal cancer were reviewed. At the levels for therapeutic impact, patient outcome and health economics, all clinical applications were reviewed. The description of the methodology is thus divided into two broad sections for:

- EUS in gastro-oesophageal cancer
  – staging performance
  – staging impact
- EUS in any clinical application
  – therapeutic impact
  – patient outcome
  – health economics.

3.1 EUS in gastro-oesophageal cancer – staging performance

3.1.1 Search strategy

The following electronic databases were searched and the search strategies are given in appendix 1:

- MEDLINE
- Bath Information and Data Services (BIDS) – Institute of Scientific Information (ISI)
- EMBASE
- Cochrane Library
- Inside Information Plus, British Library[14]
- FirstSearch, Online Computer Library Centre (OCLC).[15]

Unlike MEDLINE, BIDS does not classify articles into subject categories, but it does provide similar Boolean and text word capabilities. The BIDS archive extends from 1981 to the present day; however, an advantage over MEDLINE is that it is updated daily. BIDS also includes selective conference proceedings and abstracts, a service not supplied by MEDLINE. There is substantial overlap between these two databases; in addition, a small proportion of articles are unique to them individually. A comprehensive comparison of MEDLINE and BIDS search strategies identifies subtle differences between Boolean commands and also incompatibility of specialised search commands of both systems. Hence, separate search strategies were compiled for these two databases. Both MEDLINE and BIDS were searched from 1981 to the end of 1996. As there are sometimes delays before MEDLINE updates with new publications, the search strategy was re-run in October 1997 to ensure all references were up to date. The EMBASE search, performed by a library professional, was designed to have higher precision and therefore lower recall, to limit the number of inappropriate retrievals.

The two remaining resources became available in 1997, during the period of this project; they facilitated access to otherwise inaccessible journals. The first, Inside Information Plus, supplied by the British Library, enables access, searching and ordering of a large selection of their archive. The service covers 250,000 journals, of which 20,000 can be searched down to article title level with the use of keywords, along with 16,000 conference proceedings. The archive extends back only as far as 1993, but the service is updated within 72 hours of receipt of new material. The second service is provided by the OCLC, which is a non-profit computer service and research organisation whose network and services link more than 25,000 libraries in the USA and 63 countries and territories. Using a service called FirstSearch, more than 60 databases covering 14 topic areas can be searched by using keywords. These databases include: WorldCat, a merged

electronic catalogue of libraries around the world; ArticleFirst, a catalogue of individual articles; ContentsFirst, a catalogue of journal contents divided by volume and issue; NetFirst, a catalogue of Internet-accessible resources; and ProceedingsFirst, a catalogue of conference proceedings. All these databases were searched by title and subject using the keywords included in the MEDLINE search strategy.

Handsearching

Journals that were cited by one of the main electronic databases (MEDLINE or BIDS) were not handsearched. Because high-recall search criteria were used and additional resources were searched, the impact of not undertaking this extensive task is considered to be negligible. Uncited journals were identified from: the reference lists of articles in cited journals, the ISI citations lists, browsing library catalogues and Internet websites. The references of all retrieved articles were handsearched to identify any additional studies.

The journals not cited by MEDLINE or BIDS are listed in *Table 3.1*. A large proportion of journals initially identified as needing handsearching are included in the two new services, Inside Information Plus and FirstSearch, thus the handsearching task was substantially reduced. Ultimately, no journals were handsearched. The seven journals that were not possible to search electronically were excluded for the following reasons:

- *Gastroenterological Endoscopy* was identified as non-English language, and excluded by the criteria of section 3.1.2.
- *Annual of Gastrointestinal Endoscopy, Ultrasound Quarterly, Ultrasound Annual* and *Current Oncology* were identified as review journals. It was assumed that any primary research covered would also have been published elsewhere.
- *Journal of Medical Imaging* was incorporated into the *European Journal of Radiology* for the period 1987–1989, and this journal was electronically searched on MEDLINE.
- *European Medical Ultrasonics* was discontinued in 1989 and it was not possible to gain access to its early volumes for handsearching.

Contacting authors, academic centres and manufacturers

A valuable source of identifying authors of studies either in progress or close to completion is abstracts from conference proceedings. Any abstracts that do not appear as a full article at a later date may be a casualty of publication bias. It is not always easy to discern whether a particular abstract develops into a full article because, in some cases, the title and context change as well as the authors. We therefore set out to contact all authors of abstracts unless it was clear that the study had later been published in full. Part of the questionnaire sent to authors is shown in *Figure 3.1*.

An e-mail discussion group organised by the American Endosonography Club[16] was contacted with the same request for any unpublished studies.

To investigate possible publication bias further, major academic centres of endoscopic ultrasonography were identified from the published articles. The top 17 centres in terms of the number of articles published were contacted by letter and the information regarding unpublished data shown in *Table 3.2* was requested.

Seven manufacturers were approached with a request for clinical information. Six were identified primarily for their production of CT machines, as part of our review of spiral and electron beam CT,[17] and information on EUS was also requested from these manufacturers. A letter specific to EUS was addressed to KeyMed Medical and Industrial Equipment Ltd, the UK distributors of the Olympus EUS probes.

Grey literature

The database of grey literature supplied by the British Library, the System for Information on Grey Literature (SIGLE) was searched using keywords from the MEDLINE search.

3.1.2 Inclusion criteria

Study selection was a three-stage process as illustrated below:

1. Preliminary inclusion criteria were applied manually to the returns of the electronic searches:

- published before January 1997
- not an abstract
- not a review article
- English language
- not a case report
- not an editorial
- not a letter.

2. The remaining abstracts and titles were then assessed against the inclusion criteria shown in *Table 3.3* and full copies of qualifying articles acquired.

3. The final set of inclusion criteria is shown in *Table 3.4*. These were used to select articles suitable for inclusion in the review.

TABLE 3.1 *Summary of journals not cited by MEDLINE or BIDS*

Journals not on MEDLINE or BIDS	Searched electronically	Cited by ISI	Inside Information Plus	FirstSearch
Acta Chirurgica Austriaca	✔	✗	✔	✗
Acta Endoscopica	✔	✗	✔	✔
Advanced Imaging	✔	✗	✔	✔
Annual of Gastrointestinal Endoscopy	✗	✗	✗	✗
Applied Radiology	✔	✗	✔	✔
Asian Journal of Surgery	✔	✗	✔	✗
Asian Medical Journal	✔	✗	✔	✔
Chirurgia	✔	✗	✔	✗
Clinical MRI	✔	✗	✔	✗
Contributions to Oncology	✔	✔	✗	✔
Current Gastroenterology	✔	✗	✔	✗
Current Oncology	✗	✗	✗	✗
Diagnostica	✔	✔	✔	✗
Diagnostic and Therapeutic Endoscopy	✔	✗	✔	✗
Diagnostic Imaging	✔	✗	✔	✔
Digestive Endoscopy	✔	✗	✔	✗
Diseases of the Esophagus	✔	✗	✔	✗
Emergency Radiology	✔	✗	✔	✗
European Journal of Ultrasound	✔	✗	✔	✗
European Medical Ultrasonics	✗	✗	✗	✗
Euroson Proceedings	✔	✗	✔	✗
Evidence Based Medicine	✔	✗	✔	✗
Experimental and Clinical Gastroenterology	✔	✗	✔	✗
Frontiers of Gastrointestinal Research	✔	✔	✔	✔
Gastroenterological Endoscopy	✗	✗	✗	✗
Gastroenterological Journal of Taiwan	✔	✗	✔	✗
Gastroenterological Surgery	✔	✗	✔	✗
Gastroenterology International	✔	✗	✔	✔
Hellenic Journal of Gastroenterology	✔	✗	✔	✗
Indian Journal of Radiology and Imaging	✔	✗	✔	✗
International Gastric Cancer Congress	✔	✗	✔	✗
International Journal of Gastroenterology	✔	✗	✔	✗
Journal of Clinical Nutrition and Gastroenterology	✔	✗	✔	✗
Journal of Gastroenterology and Hepatology	✔	✗	✔	✔
Journal of the Japan Society for Cancer Therapy	✔	✔	✔	✗
Journal of Medical Imaging	✗	✗	✗	✗
Journal of Tokyo Womens Medical College	✔	✗	✔	✗
Practical Gastroenterology	✔	✗	✔	✔
Stomach and Intestine	✔	✗	✔	✗
Surgical Research Communications	✔	✗	✔	✔
Ultrasound Annual	✗	✗	✗	✗
Ultrasound Quarterly	✗	✗	✗	✗
World Congress – International College of Surgeons	✔	✗	✔	✗

❑ The study has been accepted for publication in a peer-reviewed journal.
The reference is:

If you are able to send us a reprint we would be most grateful.

❑ The study has been submitted for publication but rejected.

❑ The study is not yet completed.

FIGURE 3.1 *Part of a questionnaire sent to authors of abstracts published in conference proceedings*

TABLE 3.2 *Part of a questionnaire sent to major academic centres of endoscopic ultrasonography to identify publication bias*

	No. studies
Study was presented at a meeting but the abstract did not appear in an indexed journal	
Study was submitted for presentation at one or more meetings but was rejected	
Study was not submitted for presentation at a meeting	
Study was submitted for publication but rejected	
Study was not submitted for publication	

TABLE 3.3 *Subject-specific inclusion criteria*

Inclusion criterion	
Anatomical location	Oesophagus, stomach or cardia
Type of disease	Squamous cell carcinoma or adenocarcinoma
Use of EUS	Preoperative staging Miniprobes FNA
Type of study	Staging performance

TABLE 3.4 *Reliability and validity inclusion criteria*

Inclusion criterion	**Value**
No. of patients	> 10
Adequate gold standard	Pathology/histology
Sufficient raw data presented	To enable completion of 2 x 2 contingency table
Sufficient study information	At least basic details of study undertaken
Original patient data set	The most recent report of a patient group used

For studies that did not meet the inclusion criterion covering raw data, but satisfied the others, the corresponding author was contacted and asked to provide the additional data necessary to complete the 2×2 table.

3.1.3 Assessment of relevance and validity of primary studies

The Centre for Reviews and Dissemination Report 4[18] recommends grading primary studies into a hierarchy according to their design. Level I includes well-designed RCTs, level II includes both prospective and retrospective controlled trials, while level III covers comparisons lacking controls. Level IV is opinion-based evidence. Because controlled studies are not found in our topic area, meaning that all our evidence falls in level III, this hierarchy proved to be inapplicable to this review. Instead, it was decided to begin our assessment of validity at a lower level, by determining the possible presence in the study design of bias likely to threaten validity. We drew up a list of 20 potential biases, which are shown in *Table 3.5* and described by Kelly *et al.*[6]

Identifying the presence of bias

A checklist approach was required, but we found that none of those published at the time we began our review was suited to the application.

TABLE 3.5 *Potential biases in diagnostic imaging studies*[6]

Subjects			**Study**		**Interpretation**
Patient selection			***Application of the gold standard***		***Independence of interpretations***
Referral bias	*Patient filtering*	*Patient cohort*	Verification bias		Diagnostic review
Centripetal	Diagnostic safety	Spectrum	Work-up bias		Test review
Popularity	Co-intervention	Population	Incorporation bias		Comparator review
Diagnostic access					Clinical review
			Measurement of results		
			Disease progression	*Withdrawal bias*	*Observer variability*
				Indeterminate results	Intrinsic interobserver
				Loss to follow-up	Extrinsic interobserver
					Intraobserver
Main effect external validity			**Main effect internal validity**		**Main effect internal validity**

Those designed for RCTs were inapplicable, those designed for observational studies were best for controlled trials of treatment, and even those for diagnostic tests[19–21] were not as generic as we wished. For example, in medical imaging a test may be used for purposes other than to differentiate between diseased and disease-free individuals. Tests may be used for staging disease, as part of a diagnostic work-up, or to guide other procedures. We required a checklist that was generic enough to be adapted quickly to suit both pure diagnostic applications and the other possibilities. Additionally, those features of the conduct of the studies that might vary between them had to be noted to allow proper comparison. We have chosen to call this category 'factors', and our checklist has a separate section to note information relating to the equipment used and the imaging protocol. The checklist is a two-part document: the questions and essential guidelines. A checklist for EUS is shown in appendix 2. It comprises 30 questions divided into four major sections. The first section covers the focus and basic details of the article, and the remaining three sections cover biases due to patient selection and to study conduct, and independence of interpretation biases. The checklist was designed to assess individual study quality by containing specific questions applicable to each of the potential biases, while maintaining a broad applicability over all diagnostic performance studies. In order that the answers to the questions should be reproducible and objective, very specific guidelines are required. The guidelines may require slight modification for different clinical applications.

Our checklist is compatible with the suggestions of the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests.[22] The checklist covers very similar points, but is presented as a series of questions.

Assessment of the checklist's interobserver reliability was performed on two separate occasions. First, when the checklist was in its preliminary form, seven of the review panel members used it independently to evaluate three arbitrarily chosen EUS studies that had not been discussed beforehand. The results were collected and compared with the results recorded by the main reviewer. All discrepancies were analysed by the group and refinements were made according to consensus decisions. It was discovered from this process that agreement was high when the information was clearly presented in the study report, but poor agreement was found when information was vague or missing. This emphasised the importance of an objective checklist that excludes any judgement from the reviewer.

The second interobserver test was undertaken towards the end of the review, when the checklist was fully developed and in use. For this trial, the participants were not from the review panel. They were clinical and research scientists who had no prior knowledge of systematic reviewing or the biases involved. They received no training in the use of the checklist before the trial. Four people participated and completed the checklist independently for one EUS study. They were also asked to record an estimate of the length of time it took them to complete the checklist. Again the results were compared with those of the main reviewer and discrepancies were discussed individually. On average the agreement was good: seven questions attained 100% agreement; 11 questions attained 80% agreement; five questions attained 60% agreement; and the remaining three questions were found to be misinterpreted due to unfamiliarity with some of the terminology used or from overlooking information. When comparing each observer's answers with those of the reviewer, the calculated kappa statistic for the whole checklist ranged from 0.26 to 0.67 (average 0.48), indicating a fair to moderate performance. The time taken to complete the checklist ranged from 16 to 45 minutes (average 31.5). From this experience is was concluded that the checklist was sufficiently reproducible and objective, and reasonably simple to complete. The participants had received no training before completing the checklist and they still achieved reasonable comparability. The results highlighted areas that should be covered before new users could begin reviewing.

Ranking study validity

After completing the checklist, we had initially hoped to be able to rank the biases in order of significance (in a manner similar to that described by Mulrow[19]) and develop a numerical scoring scheme that would allow the objective ranking of studies by validity.[23] Investigation of this approach is described in appendix 3. The results showed that this approach introduced unwanted subjectivity and was abandoned because no consensus could be reached on the relative importance of the biases. Even if an unweighted combination is considered there are difficulties. Studies in a given subject tend to have properties in common, perhaps dictated by the clinical application area, which means that they share common faults in study conduct, data interpretation or patient selection, and will not be differentiated in such a scheme.

A further difficulty arose because of a widespread lack of reporting of study design in the medical imaging literature. In common with most authors, we chose to rate a study as having the risk of bias if the information required to determine whether or not there is a risk of bias was not given. A very high proportion of studies fell into this category and it was clear that the final review would exclude potentially valid results because of this. Instead, all studies meeting the inclusion criteria *(Table 3.4)* were included in a statistical analysis (described in section 3.1.6) designed to determine if the results of the study were related to the likely presence of one or more biases.

3.1.4 Data extraction

Data were extracted from the selected studies by the main reviewer. As well as the information required to complete the factor and bias checklists, the staging performance results were also extracted (appendix 4). This included results for staging tumours, lymph nodes, metastases and an overall grouped TNM stage, according to the 1987 TNM system,[2] for both EUS and any comparative results for CT. To facilitate this process and to make it accurate and reliable, standard tables were used. These tables are shown in *Figure 3.2a* and are equivalent to traditional 2×2 contingency tables, which compare the test result with that of the gold standard. For tumour staging there are initially four possibilities: T1, T2, T3 and T4. These are then combined into two groups to allow application to a 2×2 table. The grouping of T1 and T2 versus T3 and T4 was chosen as that likely to be of the most clinical significance. For lymph nodes, depending on the anatomical site, there are three possibilities, N0, N1 and N2. For oesophageal lymph nodes only N0 and N1 are applicable; therefore for the stomach N1 and N2 were combined together to allow the completion of the 2×2 table.

For the grouped TNM stage results *(Figure 3.2b),* two systems are identified, one for each anatomical location, and the combination of T, N and M for each stage is shown.

Authors of studies that did not supply sufficient information to complete the 2×2 contingency table were contacted, requesting the necessary information.

3.1.5 Quantitative data synthesis

The results of each primary study were expressed using the summary statistics: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and the odds ratio (OR) *(Figure 3.3)*. The results for each study are shown in appendix 4. It is possible that all of these statistics were dependent on the specific threshold between positive and negative used in each study to

Tumour staging

	Gold standard				
EUS	T1	T2	T3	T4	Total
T1					
T2					
T3					
T4					
Total					

	Gold standard		
EUS	T1/T2	T3/T4	Total
T1/T2			
T3/T4			
Total			

Lymph node staging

	Gold standard			
EUS	N0	N1	N2	Total
N0				
N1				
N2				
Total				

	Gold standard		
EUS	N1/N2	N0	Total
N1/N2			
N0			
Total			

Staging of metastases

	Gold standard		
EUS	M1	M0	Total
M1			
M0			
Total			

FIGURE 3.2a *Result tables for EUS – separate T, N and M staging*

Grouped TNM staging

Oesophagus

	Histological stage					
EUS	I	IIA	IIB	III	IV	Total
I						
IIA						
IIB						
III						
IV						
Total						

Stage I: T1 N0 M0
Stage IIA: T2 N0 M0 or T3 N0 M0
Stage IIB: T1 N1 M0 or T2 N1 M0
Stage III: T3 N1 M0 or T4 any N M0
Stage IV: any T any N M1

Stomach

	Histological stage						
EUS	IA	IB	II	IIIA	IIIB	IV	Total
IA							
IB							
II							
IIIA							
IIIB							
IV							
Total							

Stage IA: T1 N0 M0
Stage IB: T1 N1 M0 or T2 N0 M0
Stage II: T1 N2 M0 or T2 N1 M0 or T3 N0 M0
Stage IIIA: T2 N2 M0 or T3 N1 M0 or T4 N0 M0
Stage IIIB: T3 N2 M0 or T4 N1 M0
Stage IV: T4 N2 M0 or any T any N M1

FIGURE 3.2b *Result tables for EUS – grouped TNM staging*

interpret the test results. A range of values is expected for each statistic, illustrating the differing thresholds between studies. This is analogous to receiver operator characteristic (ROC) methodology,[24] where the existence of varying thresholds is used to plot a curve representing the performance of the test regardless of the particular observer involved or their chosen threshold. It is then possible to select a threshold between positive and negative that gives the desired balance of sensitivity and specificity. In this review, the independent

	Gold standard		
Test	Positive	Negative	Total
Positive	TP	FP	TP + FP
Negative	FN	TN	FN + TN
Total	TP + FN	FP + TN	N

$$Sensitivity = \frac{TP}{TP + FN} \qquad Specificity = \frac{TN}{TN + FP}$$

$$PPV = \frac{TP}{TP + FP} \qquad NPV = \frac{TN}{TN + FN}$$

$$Accuracy = \frac{TP + TN}{N} \qquad OR = \frac{TPR/(1 - TPR)}{FPR/(1 - FPR)}$$

FIGURE 3.3 *2 x 2 Contingency table and equations for expressing staging performance (TP = true-positive, TN = true-negative, FP = false-positive, FN = false-negative, N = TP + TN + FP + FN, PPV = positive-predictive value, NPV = negative-predictive value, OR = odds ratio, TPR = true-positive rate = sensitivity, FPR = false-positive rate = 1 – specificity)*

trials were combined using the methodology developed by Moses *et al.*[25] and Irwig *et al.*,[26,27] which expands on the principles of ROC analysis. The overall procedure is outlined in *Table 3.6*.

Stage 1: ROC scatterplot to visualise range of results from the primary studies

An ROC scatterplot was made to illustrate the range of results for the tests. It was used to determine whether any of the individual results lay outside the area of decision making, which has been defined as sensitivity and specificity both greater than 50%.[25] Two analyses were performed; the first included only those results within that clinically relevant range and the second used the full set of results.[26]

Stage 2: Summary ROC curve to estimate a best fit to the data and remove the effect of possible relations between the results and the threshold used for classifying a study as positive

A summary ROC (SROC) curve (vertical axis D versus horizontal axis S) was fitted according to the linear model shown in equation 3.1, in order to remove the effect of possible relations between the results and the threshold used within a study for classifying results as positive. To avoid problems from missing points due to zero cells of the 2×2 contingency table (i.e. a test with either zero or 100% sensitivity or specificity), the value in each cell was increased by 0.5.[25] Two models[25] were used to fit the line on the SROC curve: the equally weighted least squares (EWLS) method and a robust-resistant (RR) method.

The intercept, A, of the model is the estimated log OR when sensitivity equals specificity (S = 0). The gradient, B, provides an estimate of the extent to which the log OR is dependent on the threshold used. If B is zero, the log OR is independent of threshold, and test accuracy for each primary study is summarised by a common OR, given by the intercept.

$$D = A + BS \qquad \text{Equation 3.1}$$

where:

$D = \text{logit}(TPR) - \text{logit}(FPR)$

$S = \text{logit}(TPR) + \text{logit}(FPR)$

and

$\text{logit}(TPR) = \log\{TPR/(1 - TPR)\}$

$\text{logit}(FPR) = \log\{FPR/(1 - FPR)\}$

A = intercept

B = gradient

TPR = true-positive rate

FPR = false-positive rate

TABLE 3.6 *Summary of data synthesis procedure*

Stage	Plot	Reason
Stage 1	ROC scatterplot	To visualise the range of results from the primary studies Variations are assumed to be due to the use of differing thresholds for defining positivity
Stage 2	SROC curve with axes D and S	Straight lines fitted accordingly, to estimate a best fit to the data and remove the effect of possible relations between the results and the threshold used for classifying a study as positive
Stage 3	SROC curve with conventional axes TPR and FPR	To present the combination of results from primary studies as a single ROC curve

ROC, receiver operator characteristic; SROC, summary ROC curve; D, vertical axis of SROC; S, horizontal axis of SROC; TPR, true-positive rate; FPR, false-positive rate

Stage 3: ROC curve to present combination of results from primary studies

A conventional ROC curve, with TPR and FPR axes, was plotted to summarise the combined results. Equation 3.2 was used to convert back to the conventional axes, with substitution of the gradient and intercept values calculated from the model of Stage 2, providing the gradient calculated was non-zero.

$$TPR = \left[1 + e^{-A/(1-B)} \left(\frac{1 - FPR}{FPR} \right)^{\frac{(1+B)}{(1-B)}} \right]^{-1} \quad \text{Equation 3.2}$$

To give a single value to summarise this ROC curve, the point on the curve where sensitivity is equal to specificity was used: denoted Q*. This value was obtained from the intercept of the ROC curve and a line plotting sensitivity equals specificity *(Figure 3.4)*. Q* is the most appropriate summary statistic for EUS and represents the optimum performance for the following reasons. Due to the dichotomy chosen for tumour staging (section 3.1.4), T1 or T2 is analogous to a positive diagnosis in a conventional 2 × 2 table, and therefore T3 or T4 is analogous to a negative diagnosis. This implies that sensitivity is a measure of the ability of EUS to stage T1/T2 correctly and not overstage tumours as T3/T4. Conversely, specificity is a measure of the ability of EUS to stage T3/T4 correctly and not understage tumours as T1/T2. Neither understaging nor overstaging can be assumed to have more or less impact than the other: understaging tumours will result in surgical operations that are unnecessary, and overstaging will result in palliative or non-surgical treatments when resection may have been possible. The most appropriate threshold is one that minimises both understaging and overstaging (i.e. Q*, which balances sensitivity and specificity).

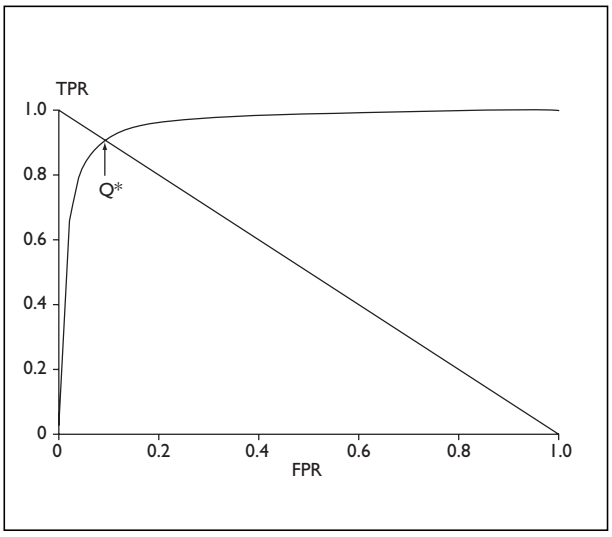


FIGURE 3.4 *Illustration of summary estimate, Q**

3.1.6 Differences between studies

Regression techniques were used to analyse the influence of differences between the studies on the summary results. The covariates for the analysis were the biases and factors described in section 3.1.3. *Table 3.7* lists those included for tumour and lymph node evaluation. The lists differ and not all those initially identified are included in *Table 3.7* because the studies had similar gaps in information and, also, some biases were present (or absent) in all articles. The blinding biases represent the four biases labelled under 'Independence of interpretations' in *Table 3.5*. If any attempt to perform blinding was reported, this combined bias was regarded to be absent.

TABLE 3.7 *Factors and biases included in the multivariate analysis of differences between studies*

Topic	Biases	Factors
Tumour	Verification bias Disease progression bias Withdrawal bias Blinding biases	Gold standard Model Frequency Stenosis Year Patient number
Lymph nodes	Verification bias Disease progression bias Withdrawal bias Blinding biases	Gold standard Model Frequency Size Year Patient number

For this part of the analysis, the EWLS plot was used and the factors or biases were analysed by incorporating them as a multivariate extension into equation 3.1.[25] Using this methodology, the SROC curve is divided into two separate plots for each variable assessed; *Figure 3.5* illustrates the principles. The gradients of each of the two separate plots are made equal. The fit of this gradient is dependent on the EWLS plots of each individual set, the resulting gradient being a compromise between the two. If the variable is found to be significant in the regression analysis, then each of the two sets will have a significantly different intercept. Compared with the intercept of the total data set, one set will have a higher intercept and the other a lower intercept. Thus the range

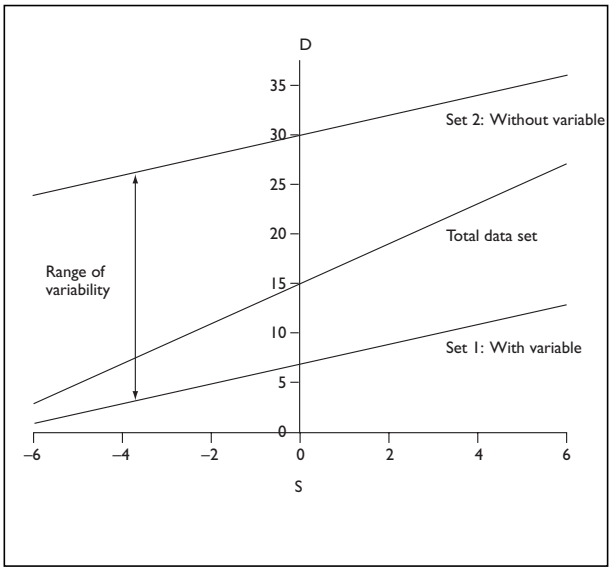


FIGURE 3.5 *Illustration of the regression analysis on an SROC curve*

between the two intercepts represents the likely variance of the overall result.

If no factors or biases were found to be significant when they were all combined into the regression analysis, then their individual influence was assessed.

3.1.7 Reproducibility and learning curve

The database of studies retrieved from the main staging performance search was examined for those addressing reproducibility and learning curve issues. Summaries of relevant studies were prepared.

3.2 EUS in gastro-oesophageal cancer – staging impact

3.2.1 Search strategy

The same electronic databases were searched as for the staging performance search. The additional search strategy for staging impact, therapeutic impact and patient outcome is shown in appendix 1. Any relevant articles applicable to this level were also identified and retrieved during the reviewing process of the staging performance articles. In the absence of studies designed intentionally to evaluate staging impact, those using comparative technologies in a staging performance study design were sought for use as a secondary standard.

3.2.2 Inclusion criteria

Preliminary inclusion criteria were automatically applied to the returns of the electronic searches:

- published before January 1997
- not an abstract
- not a review article
- English language
- not a case report
- not an editorial
- not a letter.

The abstracts of these articles were then read and a decision was made about the relevance of each study in terms of the applicability to EUS, the fulfilment of the above criteria, and the following exclusion criteria:

- keyword used in a different context from that intended in our search
- fewer than or equal to ten patients
- non-human study.

Full copies of qualifying papers were acquired and checked against the criteria already set. Further exclusions were made as necessary.

The decision of whether or not to include the article was based on simple criteria, assessed from reading the full paper. Because the level of information was poor, these basic inclusion criteria were all that were used for staging-impact studies:

- original data reported (not just qualitative discussion)
- comparative study of staging with and without EUS
- link between use of EUS and changes in staging decisions or confidence in those decisions.

3.2.3 Assessment of relevance and validity of primary studies

The criteria for a valid study designed specifically to evaluate staging impact have been discussed in chapter 1. No formal assessment was performed owing to the paucity of data. The limitations of studies identified are detailed in section 8.2.

The generic bias checklist was applied to studies that compared EUS with other modalities. In the event, the only technology compared with the performance of EUS was incremental CT. For the bias checklist, no alterations were necessary to improve its applicability to CT, and, for the factors checklist, only the technical section required moderate alteration.

3.2.4 Data extraction

The available data on the performance of incremental CT were extracted by the same method as for EUS, as described in section 3.1.4 using the tables of *Figure 3.6*.

	Gold standard				
CT	T1	T2	T3	T4	Total
T1/T2					
T3					
T4					
Total					

	Gold standard		
CT	T1/T2	T3/T4	Total
T1/T2			
T3/T4			
Total			

FIGURE 3.6 *Data extraction tables for CT*

A significant difference with incremental CT compared with EUS is that it cannot differentiate between all of the four T stages. In the literature analysed there were two methodologies used to account for this. In five of eight articles using CT and EUS, it was claimed that CT could not differentiate between T1 and T2; hence these two separate stages were combined together. In the remaining three of eight articles, it was claimed that CT could not differentiate between T2 and T3; hence these two stages were combined. In the first example we could compare directly the results of EUS and CT because our chosen dichotomy was T1 and T2 versus T3 and T4. However, for the CT studies that combined T2 and T3, no comparison was made using the results of these studies because this did not fit into our dichotomy.

3.2.5 Data synthesis

Quantitative data synthesis as described in section 3.1.5 was planned but not performed. The reasons for this are discussed in section 8.2.

3.3 EUS in any clinical application – therapeutic impact

3.3.1 Search strategy

The same electronic databases were searched as for staging performance. The additional search strategy for staging impact, therapeutic impact and patient outcome is shown in appendix 1. Any relevant articles applicable to this level were also identified and retrieved during the reviewing process of the staging performance articles.

3.3.2 Inclusion criteria

Preliminary inclusion criteria were automatically applied to the returns of the electronic searches:

- published before January 1997
- not an abstract
- not a review article
- English language
- not a case report
- not an editorial
- not a letter
- neither the anatomical area nor the clinical application of EUS was used as an exclusion criterion.

The abstracts of these articles were then read and a decision was made about the relevance of each study in terms of the applicability to EUS, the fulfilment of the above criteria, and the following exclusion criteria:

- keyword used in a different context from that intended in our search
- less than or equal to ten patients
- non-human study.

Full copies of qualifying articles were acquired and checked against the criteria already set. Further exclusions were made as necessary.

The decision of whether or not to include an article was based on simple criteria, assessed from reading the full text. Because the level of information was poor, these basic criteria were all that were used for therapeutic-impact studies:

- original data reported (not just qualitative discussion)
- comparative study of staging with and without EUS
- link between use of EUS and changes in therapeutic decisions or confidence in those decisions.

3.3.3 Assessment of relevance and validity of primary studies

Criteria for a valid study designed specifically to evaluate therapeutic impact have been discussed in chapter 1. No formal assessment was performed owing to the paucity of data. The limitations of studies identified are detailed in section 8.3.

3.3.4 Data extraction

Descriptive summaries were written.

3.3.5 Data synthesis

Quantitative data synthesis was not applicable at the therapeutic impact level. Narrative synthesis was planned if sufficiently similar studies were included.

3.4 EUS in any clinical application – patient outcome

3.4.1 Search strategy

The same electronic databases were searched as for the staging performance search. The additional search strategy for staging impact, therapeutic impact and patient outcome is shown in appendix 1. Any relevant articles applicable to this level were also identified and retrieved during the reviewing process of the staging performance articles. The culmination of positive or negative diagnostic and therapeutic impact may be a change in patient outcome. Studies not following strict methodologies to evaluate patient outcome but supplying limited information regarding patient survival were identified but were not further assessed.

3.4.2 Inclusion criteria

Preliminary inclusion criteria were automatically applied to the returns of the electronic searches:

- published before January 1997
- not an abstract
- not a review article
- English language
- not a case report
- not an editorial
- not a letter
- neither the anatomical area nor the clinical application of EUS was used as an exclusion criterion.

The abstracts of these articles were then read and a decision made about the relevance of each study in terms of the applicability to EUS, the fulfilment of the above criteria, and the following exclusion criteria:

- keyword used in a different context from that intended in our search
- less than or equal to ten patients
- non-human study.

Full copies of qualifying articles were acquired and checked against the criteria already set. Further exclusions were made as necessary.

The decision of whether or not to include an article was based on simple criteria, assessed from reading the full article. Because the level of information was poor, these basic criteria were all that were used for patient outcome studies:

- original data reported (not just qualitative discussion)
- comparative study of staging with and without EUS
- link between use of EUS and patient outcome.

3.4.3 Assessment of relevance and validity of primary studies

As in many other aspects, the imaging literature reviewed was lacking in good outcome studies. As a consequence, a strict checklist of criteria was not applied and all studies that promised any outcome data were included. The criteria for a valid study designed specifically to evaluate patient outcome have been discussed in chapter 1. No formal assessment was performed owing to the paucity of data. Limitations of the studies identified are detailed in section 8.4.

3.4.4 Data extraction

Descriptive summaries were written.

3.4.5 Data synthesis

Quantitative data synthesis was not applicable at the patient outcome level. Narrative synthesis was planned if sufficiently similar studies were included.

3.5 EUS in any clinical application – health economics

3.5.1 Search strategy

The initial electronic search strategy for health economics studies was based on the combination of search terms for economics, endoscopic ultrasound and the specific clinical problem (the staging of gastro-oesophageal tumours). It became rapidly apparent that this combination was too restrictive given the relatively small number of studies on the economics of imaging in general. A broader search strategy was then adopted using search terms for economics and endoscopic ultrasound in general.

As well as the *Index Medicus* medical subject heading (MeSH) category, 'economics', several relevant economic textword indicators were used individually to ensure a search with high sensitivity. Other terms frequently (but not exclusively) used in economic studies (benefit, impact, management, outcome and utility) were limited to any combination of two to balance between retrieving all

relevant studies and minimising the identification of inappropriate articles.

The search terms used in MEDLINE and BIDS are given in appendix 1.

Handsearching was undertaken on the bibliographies of articles identified in the electronic search, and on selected health articles identified in the electronic search and in selected health economics journals. The following journals were handsearched:

- *International Journal of Technology Assessment in Health Care*
- *Health Economics*
- *Health Policy*
- *Social Science in Medicine.*

Abstracts from the 1995, 1996 and 1997 International Society of Technology Assessment in Health Care (ISTAHC) conferences were also handsearched to identify studies that were under way.

3.5.2 Inclusion criteria

Preliminary inclusion criteria were applied to the returns of the electronic searches:

- published before January 1997
- not an astract
- not a review article
- English language
- not a case report
- not an editorial
- not a letter
- neither the anatomical area nor the clinical application of EUS was used as an exclusion criterion.

The abstracts of these articles were then read and a decision was made about the relevance of each study in terms of the applicability to EUS and the fulfilment of the above criteria.

Full copies of qualifying articles were acquired and checked against the criteria already set. Further exclusions were made as necessary.

The decision of whether or not to include an article was based solely on economic information being reported.

3.5.3 Assessment of relevance and validity of primary studies

The criteria for judging the quality of economic studies in health care differ from those used to assess studies of diagnostic accuracy. The focus of economic evaluation is on resource use and benefits to patients, which may be realised in a routine healthcare delivery situation. In other words, the external validity of studies is more important than the internal validity, which means that RCTs have drawbacks as vehicles for economic evaluation. Nevertheless, the criteria for judging economic studies can still be grouped into four main categories: study design, data collection, analysis, and interpretation of results.

Economic evaluations of healthcare technologies have been regularly undertaken for over 25 years. Agreement on the most appropriate methods has been facilitated by practical experience and the refinement of economic techniques. As a result there is no shortage of guidelines and checklists to assist the reader of economic studies. Williams[28] sets out the fundamentals, which have subsequently been elaborated by the Department of Biostatistics, McMaster University[29] and Drummond and colleagues.[30] In 1996 the *British Medical Journal* published a set of guidelines for use by reviewers of economic submissions to the Journal.[10] Condensed lists of key factors have been used by some authors in empirical studies of the quality of economic evaluations found in the clinical literature.[31,32] Adaptations of these published guidelines have been used in other recent studies of economic evaluation of diagnostic imaging.[33]

The initial intention was to attempt something similar in this review. However, it became apparent early on in the project that the quality of the economic analyses in the studies located was so poor that the use of a long checklist to assess quality was redundant. The proposed checklist, which was designed but not used, is shown in appendix 5.

To classify studies into those which:

- had adequate economic analyses
- had poor economic analyses
- could not legitimately be called economic analyses,

the following four criteria were identified as sufficient.

- Was the type of economic analysis correctly chosen and designed?
- Was the outcome indicator appropriate?
- Was the cost analysis correctly conducted?
- Was sensitivity analysis carried out?

3.5.4 Data extraction
Descriptive summaries were written.

3.5.5 Data synthesis
Data synthesis was not applicable at the health economics level.

Conclusion

In this chapter we have described the methodology of our review, dividing the studies reviewed into five levels of the Fineberg framework. In the next chapter details of those studies satisfying the inclusion criteria are presented, as outlined in the '3.n.4 Data extraction' sections of the current chapter. *Table 3.8* summarises where in this report the results of applying the various parts of the methodology are presented.

TABLE 3.8 *Summary of presentation of results: 'n' represents the subsection numbering according to the Fineberg framework, from 1 for staging performance to 5 for health economics*

Methodological section	Description	Results chapter(s)
3.n.1	Search strategy	4.0
3.n.2	Inclusion criteria	4.n for included, 5.n for excluded
3.n.3	Assessment of relevance and validity of primary studies	7
3.n.4	Data extraction	4
3.n.5	Data synthesis	6
3.n.6	Differences between studies	7

Chapter 4

Details of studies included in the review

In this chapter, the first section covers the outcome of the search methodology for all levels of the Fineberg framework. It is followed by details of the studies included in the review at each level of the framework. In order to maintain the subsection numbering established in chapter 3 for the individual levels of the framework, this first section concerned with the search methodology has been numbered 4.0. The details of the studies included in the review at each level of the framework then appear in sections 4.1 to 4.5, which correspond to sections 3.1 to 3.5.

4.0 Detailed analysis of search methodology

4.0.1 Staging performance

Electronic searches

The combination of the MEDLINE and BIDS searches retrieved 4405 articles. Other searches, shown in *Table 4.1,* found an additional 48 articles, bringing the total to 4453.

On applying the initial set of exclusion criteria shown in *Table 4.2,* 2107 studies were excluded and 2346 remained. The classifications shown are not exclusive; for example, a review article could also be counted as a non-English language article.

TABLE 4.1 *Number of articles retrieved from each resource*

Search resource	No. articles
MEDLINE and BIDS	4405
EMBASE only	2
Reference lists only	21
Inside Information Plus only	7
FirstSearch only	8
EMBASE and reference lists	3
EMBASE and Inside Information Plus	2
EMBASE and FirstSearch	1
Inside Information Plus and FirstSearch	4
Any database	4453

TABLE 4.2 *Initial exclusion criteria*

Exclusion criterion	No. articles
Published in 1997	89
Abstract	744
Review article	487
Non-English language	959
Case report	32
Editorial	38
Letter	55
Unique articles excluded	2107

When the subject-specific inclusion criteria of *Tables 3.3* and *4.3* were applied, using the information provided in the articles' abstracts, the number of studies was reduced from 2346 to 186. These were either deemed to be of potential value or required reading of the full articles to decide their applicability. The articles were retrieved and their full text reviewed. This process identified a further 67 articles that did not meet the inclusion criteria already set out, bringing the number to 119. From these 119 studies, 83 were excluded for the reasons highlighted in chapter 5, leaving the 36 studies included in the review.[1–36] Of these, seven (one of which is also applicable to staging performance) studied FNA; three concerned miniprobes. Twenty-seven articles evaluated the staging performance of EUS: ten of the oesophagus; ten of the stomach; three of both the stomach and the oesophagus;

TABLE 4.3 *Inclusion criteria based on focus of review*

Inclusion criterion	Chosen value
Anatomical location	Oesophagus, stomach or cardia
Type of disease	Squamous cell carcinoma or adenocarcinoma
Use of EUS	Preoperative staging Miniprobes FNA
Type of study	Staging performance

TABLE 4.4 *Source of studies included in the review*

Resource	No. studies included in the review
MEDLINE or BIDS	33
Inside Information Plus	1
FirstSearch	1
Reference lists	1

and four of the cardia (one including only tumours at the cardia, one including tumours at the cardia and in the oesophagus, and two including tumours of the oesophagus and the gastro-oesophageal junction). *Table 4.4* shows from which search resource these 36 studies were retrieved.

Contacting authors of abstracts

Of 744 EUS abstracts identified from searching BIDS, 249 were highlighted as applicable. Of these, 172 authors were contacted. *Figure 4.1* illustrates the outcome. Of the seven studies corresponding to the first two responses, none was a study that our search had not identified.

E-mail discussion group

This approach achieved four replies. One was an ongoing trial, another reply provided two studies in press, and one detailed not a study but an Internet website under the supervision of the European Society of Gastrointestinal Endoscopy.[34] This is an excellent site, which systematically organises the available literature on EUS into topic categories in a hierarchical manner, but no additional articles were found. The final response from the e-mail list detailed the only case we identified of a study that was repeatedly not accepted for publication. The topic was not applicable to our review.

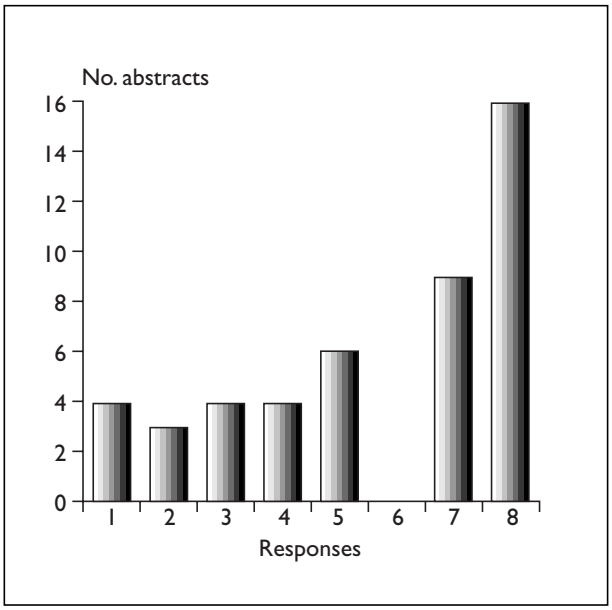


FIGURE 4.1 *Results of contacting authors of abstracts (Responses: 1, reprint supplied; 2, reference supplied; 3, published in 1997; 4, in press; 5, submitted for publication; 6, rejected; 7, not completed yet; 8, returned to sender)*

Major academic centres

Only one centre replied. The response was zero for all categories.

Manufacturers

Three replies were obtained, from Siemens, Toshiba and Philips. Toshiba gave no information, Siemens supplied information on CT but had none on EUS, Philips supplied information on both topics. Two articles on EUS were provided by Philips but they were not applicable to our review.

Grey literature

A few non-English language PhD theses and medical evaluation reports were retrieved from SIGLE but there was nothing that met the inclusion criteria.

Requesting additional data

Eight authors of studies that did not supply sufficient information to complete the 2×2 contingency table were contacted, requesting the necessary information. One reply was received, which enabled the study by Greenberg *et al.*[13] to be included in the review.

4.0.2 Staging impact

Eight studies comparing EUS with incremental CT were found. In addition, studies designed to evaluate staging impact were sought during the search for therapeutic impact and patient outcome studies, which is described next.

4.0.3 and 4.0.4 Therapeutic impact and patient outcome

Unlike the other searches, the exclusion of review articles, abstracts and non-English language articles was performed via MEDLINE and BIDS; in addition, publication dates were limited to pre-1997. This reduced the number of articles to be downloaded into the database for manual searching, but did not allow calculation of the numbers excluded. On combining MEDLINE and BIDS retrievals, 595 articles were identified. After reading the abstracts of these articles, 36 were deemed to be sufficiently applicable to the topic to require reading of the full text. The reasons for excluding these 559 studies included the criteria listed in section 3.3.2.

On reading the full articles, 14 of the 36 were excluded for one of the criteria already described, whereas 16 of them were excluded for reasons shown in chapter 5. Of the six remaining studies,[37–42] two supplied information on staging impact and on therapeutic impact, two on therapeutic impact, and two on patient outcome.

4.0.5 Health economics

From a combination of the MEDLINE and BIDS search strategies, 179 articles were retrieved that were published between 1981 and the end of 1996. Of these, 84 were not applicable to EUS, 14 were abstracts from conference proceedings and 39 were review articles; this left 42 original articles. Four further articles were identified from searching the reference lists of all retrieved articles; however, none was found to be applicable to EUS. The 42 articles were divided according to the following languages:

- English 32
- German 4
- French 2
- Japanese 1
- Russian 1
- Italian 1
- Polish 1

No additional articles were found from handsearching uncited journals.

Only two of the 32 English language articles could properly be described as economic studies,[43,44] but neither of these would score very well against any of the formal checklists.

Of the 30 articles proved unsuitable for use in the economics review, 24 were reviewed for therapeutic impact or patient outcome. The six not included in this category did not contain any reference to economics, except a brief mention in the abstract or conclusions, in what were essentially clinical studies. Some articles discussed economic issues or made assertions about cost-effectiveness without presenting any economic data. These 30 articles are listed in the reference section.[100–129]

4.1 EUS in gastro-oesophageal cancer – staging performance

4.1.1 Tumour staging – descriptive summaries

Studies that were suitable for inclusion in this review are listed in *Table 4.5* for conventional EUS probes and *Table 4.6* for miniprobes.

'Numbers of patients' refers to the numbers of T1, T2, T3 or T4 tumours for which full pathological correlation was available, but excludes those which, for whatever reason, were not surgically resected. All tumours were carcinomas, either adenocarcinoma or squamous cell. Other tumours such as lymphoma and stromal cell tumours have been excluded from the review. The majority of articles compare tumour staging according to the internationally accepted TNM system. A small minority describe the EUS findings with details regarding which layer of the bowel wall has been invaded, from which the T stage can easily be determined. Articles that failed to include information in either format could not be included.

The vast majority of studies document the use of radial scanning echoendoscopes and there are therefore very few comparative data with linear/curved array technology. Non-traversable tumours refer to tumour stenoses that could not be negotiated with the echoendoscope and did not allow a full EUS examination. The incidence of such non-traversable tumours clearly depends upon the proportion of advanced tumours included in the study, but also upon the type of echoendoscope used. The most commonly used echoendoscope is the radial scanning probe with an approximate 13 mm diameter. One study (Binmoeller *et al.*[3]) used a smaller 'blind oesophagoprobe'. This is a single 7.5 MHz frequency radial scanning wire-guided echoendoscope without viewing optics, which is able to negotiate a greater number of tumour stenoses. Some studies report increased traversability after dilatation of the tumour. Although miniprobes will negotiate all but the tightest of stenoses, their high-frequency transducers, and hence limited depth of penetration, restrict their use because a full evaluation of the tumour and locoregional lymph nodes is often not possible. Several articles report the accuracy of EUS, even when it is not possible fully to negotiate the tumour using both linear and radial equipment. It is clear from these reports that such non-traversable tumours are highly likely to be advanced (i.e. at least T3). The high-frequency miniprobes are able to discriminate up to nine layers in the bowel wall and their ability to identify subgroups of T1 tumours is discussed. This may be of benefit if local endoscopic mucosal resection of an early tumour is being considered.

Studies addressing the feasibility of FNA with EUS are described in *Table 4.7*. Only studies with relevance to gastro-oesophageal cancer masses or lymph node involvement are included.

TABLE 4.5 *Gastro-oesophageal tumour staging EUS studies*

Author and year of publication [reference]	Centre	No. patients	Site of tumour	Type of probe	Ultrasound frequencies (MHz)	Non-traversable tumours (%)
Akahoshi *et al.* 1991[1]	Fukuoka, Japan	74	St	R	7.5/12	N/S
Altorki *et al.* 1996[2]	New York, USA	53	O, C	R	7.5/12	12 (22)
Binmoeller *et al.* 1995[3]	Hamburg, Germany	38	O	R Oesophagoprobe	7.5	9/87 (10) required dilatation
Botet *et al.* 1991[4]	New York, USA	50	O	R	7.5/12	13 (26)
Botet *et al.* 1991[5]	New York, USA	50	St	R	7.5/12	N/S
Caletti *et al.* 1993[6]	Bologna, Italy	35	St	R	7.5/12	N/S
Catalano *et al.* 1994[7]	Cleveland, USA	100	O	R	7.5/12	Excluded
Dittler *et al.* 1993[9]	Munich, Germany	167	O	R	7.5/12	43 (26) after dilatation
Dittler *et al.* 1993[10]	Munich, Germany	254	St	R	7.5/12	N/S
François *et al.* 1996[11]	Nice, France	29	C	R	7.5/12	5 (17)
Greenberg *et al.* 1994[13]	Maywood, USA	20	O, J	R	7.5/12	2 (10)
Grimm *et al.* 1993[14]	Kiel, Germany	210	O = 63 St = 147	R	7.5	O = 4 (6) St = 4 (3)
Heintz *et al.* 1991[17]	Mainz, Germany	22	O	R	7.5/12	18/40 (45) excluded after dilatation
Hordijk *et al.* 1993[18]	Rotterdam, The Netherlands	39	O, J	R	7.5	15 (37)*
Hünerbein *et al.* 1996[19]	Berlin, Germany	79	O = 19 St = 60	L	7.5	O = 6 (31) St = 11 (18) C = 11/20 (55)
Manzoni 1993[20]	Verona, Italy	24	O	L	7.5	N/S
Massari *et al.* 1996[21]	Milan, Italy	65	St	R	7.5/12	2 (3)
Murata *et al.* 1988[22]	Tokyo, Japan	318	O = 172 St = 146	R	7.5/10	O = 96/269 (36) excluded
Murata *et al.* 1993[23]	Tokyo, Japan	317	O	R, L, M	7.5/12/15/20	(0–37) (probe dependent)
Perng *et al.* 1996[25]	Kaohsiung, Taiwan	67	St	R	7.5/12	N/S
Peters *et al.* 1994[26]	Los Angeles, USA	34	O	R	7.5/12	8/42 (19) excluded
Saito *et al.* 1991[27]	Tokyo, Japan	110	St	R	7.5/12	N/S
Shimizu *et al.* 1994[28]	Kyoto, Japan	128	St	R	7.5/12	N/S
Takemoto *et al.* 1986[29]	Yamaguchi, Japan	12	O	R, L	7.5	1 (8)
Tio *et al.* 1989[30]	Amsterdam, The Netherlands	68	St	R	7.5/12	N/S
Ziegler *et al.* 1991[35]	Berlin, Germany	37	O	L	7.5	7 (19)
Ziegler *et al.* 1993[36]	Berlin, Germany	108	St	R	7.5/12	2 (1.8)

O, oesophagus; C, cardia; St, stomach; J, junction; R, radial; L, linear/curved; M, miniprobe; N/S, not stated

* *This article considers the influence of tumour stenoses on the accuracy of EUS T staging and subdivides patients into 'good, difficult and impossible passage of the echoendoscope'*

TABLE 4.6 *Gastro-oesophageal miniprobe studies*

	Hasegawa et *al.* 1996[16]	**Murata et *al.* 1996[24]**	**Yanai et *al.* 1996[34]**
Centre	Nagoya, Japan	Tokyo, Japan	Ube, Japan
No. patients	22	53	45
% Male	86	87	73
Age (years)	44–78 (average 62)	49–83	43–87 (average 70)
Site of tumour	Oesophagus	Oesophagus	Stomach
No. lesions	25	53	47
Miniprobe	MP-PN15-08M	Sp101 and Sp501	Sp101
Type of probe	Radial	Linear and radial	Linear
Frequency	15 MHz	15 or 20 MHz	20 MHz
Gold standard	Histology	Histology	Histology
Conclusions	Mucosal carcinoma: Miniprobe 6/7 (86%) EUS 5/7 (71%)	Mucosal carcinoma: 16/19 (84%)	Mucosal carcinoma: 27/39 (69%)

4.1.2 Lymph node staging – descriptive summaries

Of the 27 studies that addressed the preoperative staging accuracy of EUS *(Table 4.5),* only 18 (67%) provided sufficient information on the accuracy of staging lymph nodes. Additional features of these studies to those that appear in *Table 4.5* are shown in *Table 4.8.* For gastric cancer studies, the division between malignant lymph node stages N1 and N2 is illustrated in *Table 4.8,* but for the purpose of analysis these two were combined together and compared against the benign lymph node stage N0. In *Table 4.8,* the numbers of patients are repeated because this varied in some studies from those shown in *Table 4.5.*

4.1.3 Staging of metastases – descriptive summaries

Only five studies supplied information on the staging accuracy of EUS for metastases, details of which are shown in *Table 4.9.*

4.1.4 Grouped TNM staging – descriptive summaries

Five studies provided information on the grouped TNM stage *(Table 4.10).* Two of these studies, Tio *et al.*[30] and Altorki *et al.*,[2] did not supply sufficient information to assess individually the accuracy of each stage. Altorki *et al.* did not supply any information regarding the data for staging of metastases. For the three remaining studies, only four stages were reported (i.e. no differentiation between Stages A and B).

4.1.5 Reproducibility and learning curve

There were two studies that addressed reproducibility. Both used the same general methodology involving repeated assessment of standard examinations recorded on video by an experienced operator.

[46] Catalano MF, Sivak MV, Bedford RA, Falk GW, van Stolk R, Presa F, *et al.* Observer variation and reproducibility of endoscopic ultrasonography. *Gastrointest Endosc* 1995;**41**:115–20.

Four experienced (> 50 examinations) and two inexperienced (< 20 examinations) observers reviewed videos of 50 patients who were examined by an uninvolved experienced endosonographer for non-obstructing oesophageal carcinoma. Observers were blinded to patient data and other clinical examinations. The represented stages of tumours were: T1 = 12, T2 = 11, T3 = 22 and T4 = 5. In all T stages the kappa statistic for both inter- and intraobserver variation was significantly better for experienced versus inexperienced observers (overall T: $\kappa = 0.66$ versus –0.01 interobserver; $\kappa = 0.69$ versus 0.29 intraobserver). For lymph node staging no significant difference between the two sets of observers was found (overall N: $\kappa = 0.66$ versus 0.52 interobserver; $\kappa = 0.56$ versus 0.49 intraobserver). For reproducibility of the T stages, the worst performance was for T1 lesions, which were frequently overstaged as T2.

TABLE 4.7 *Gastro-oesophageal EUS FNA studies*

	Chang *et al.* 1994[8]	**Giovannini *et al.* 1995[12]**	**Harada *et al.* 1996[15]**	**Hünerbein *et al.* 1996[19]**	**Vilmann 1996[31]**	**Vilmann *et al.* 1995[32]**	**Wiersema *et al.* 1994[33]**
Centre	Long Beach, CA, USA	Marseilles, France	Chiba, Japan	Berlin, Germany	Gentofte, Denmark	Gentofte, Denmark	Indianapolis, IN, USA
No. patients	38	141	19	40	44	33	50
No. sites	46	141	19	40	47	44	50
Nodes	✔	✔	✗	✔	✔	✔	✔
Masses	✔	✔	✔	✔	✗	✔	✔
Malignant	✔	✔	✗	✔	✔	✔	✔
Benign	✔	✔	✔	✔	✔	✔	✔
EUS probe	FG-32UA	FG-32UA	PEF-703A	FG-32UA	FG-32UA FG-36UX	FG-32UA	FG-32UA GF-UM3 GF-UM20
Type of needle	Pentax 23 gauge 4 cm GIP 22 gauge 180 cm	25 gauge 15 mm 25 gauge 40 mm 22 gauge 50 mm 22 gauge 60 mm	Endo-sonopsy needle 21 gauge 25 mm	0.8 cm 140 cm	GIP 21 gauge 170 cm	GIP 0.7 mm 160 cm	Stifcore 21 gauge 150 cm
No. passes	163	N/S	N/S	N/S	68	N/S	N/S
Passes per patient	4.3	1–5 (av. 3)	3	3–5	N/S	N/S	2–11 (av. 4)
Passes per site	3.5	1–5 (av. 3)	3	3–5	1–4 (av.1.4)	1–4 (av. 2)	2–11 (av. 4)
% Adequate specimen	91 (of 46)	89 (of 141)	84 (of 19)	100 (of 40)	83 (of 47)	84 (of 44)	N/S
% FNA diagnostic	86 (of 37)	80 (of 126)	88 (of 16)	95 (of 38)	80 (of 25)	97 (of 37)	86 (of 14)

[45] Burtin P, Napoleon B, Palazzo L, Roseau G, Souquet JC, Cales P. Interobserver agreement in endoscopic ultrasonography staging of esophageal and cardia cancer. *Gastrointest Endosc* 1996;**43**:20–4.

In this prospective, multicentre study, EUS examinations were recorded on video for 46 consecutive patients. These were assessed by five experienced (> 200 examinations) endosonographers. A subset of 28 examinations was reassessed by one observer 6 months later, both on video and by real-time evaluation. Patients presented with cancer of the cardia or oesophagus, and N staging was reported according to six anatomical groups. No information on the numbers in each T stage was provided, and stage T0 was included.

Interobserver variation for the overall T stage was fair ($\kappa = 0.48$), with stage T2 having the poorest agreement ($\kappa = 0.16$). For N staging, interobserver agreement was good for intra-abdominal lymph nodes ($\kappa = 0.73$), poor for subaortic and upper paratracheal ($\kappa = 0.22$; $\kappa = 0.31$), and fair to good for para-azygos, para-oesophageal and subcarinal lymph nodes ($\kappa = 0.49$, $\kappa = 0.55$, $\kappa = 0.57$). From the subset, the intraobserver variation depended significantly on whether the observation was real-time or via video. For real-time, the agreement was generally worse than for video (overall T: $\kappa = 0.51$ versus $\kappa = 0.91$).

One study compared the staging accuracy of a single observer during two different periods.

TABLE 4.8 *Lymph node staging studies of primary gastro-oesophageal tumours*

Author and year of publication [reference]	No. patients and site of tumour	Criteria used for lymph node malignancy	Division of N stages reported
Altorki *et al.* 1996[2]	C = 55	Echogenicity, shape, size	N0, N1
Binmoeller *et al.* 1995[3]	O = 38	Echogenicity, border, size (> 6 mm)	N0, N1
Botet *et al.* 1991[4]	O = 50	Echogenicity, shape, border	N0, N1
Botet *et al.* 1991[5]	St = 50	Echogenicity, shape	N0, N1, N2
Catalano *et al.* 1994*[7]	O = 100	Echogenicity, shape, border, homogeneity, size (> 10 mm)	N0, N1
Dittler *et al.* 1993[9]	O = 167	Echogenicity, shape, border	N0, N1
Dittler *et al.* 1993[10]	St = 254	Echogenicity, shape, border	N0, N1, N2
François *et al.* 1996[11]	C = 29	Border, size (ratio largest:smallest diameter < 2)	N0, N1, N2
Greenberg *et al.* 1994[13]	C = 16	Echogenicity, border	N0, N1
Grimm *et al.* 1993[14]	O = 62 St = 131	Echogenicity, border	O = N0, N1 St = N0, N1, N2
Heintz *et al.* 1991[17]	O = 19	Echogenicity, border, homogeneity	N0, N1
Hünerbein *et al.* 1996[19]	O = 17 St = 54	Echogenicity, shape, border	N0, N1
Massari *et al.* 1996[21]	St = 65	Echogenicity, border	N0, N1
Perng *et al.* 1996[25]	St = 69	Size (> 1 cm)	N0, N1, N2
Peters *et al.* 1994[26]	O = 34	Echogenicity, border	N0, N1, N2
Tio *et al.* 1989[30]	St = 72	Echogenicity, border	N0, N1, N2
Ziegler *et al.* 1991[35]	O = 37	Echogenicity, border	N0, N1
Ziegler *et al.* 1993[36]	St = 108	Echogenicity, border	N0, N1, N2

* *This study was designed to investigate the specific criteria for predicting lymph node malignancy*

[47] Fockens P, van den Brande JHM, van Dullemen HM, van Lanschot JJB, Tytgat NJT. Endosonographic T-staging of esophageal carcinoma: a learning curve. *Gastrointest Endosc* 1996;**44**:58–62.

A total of 231 endosonographs were performed for oesophageal malignancies by one endosonographer. After completing an 8-week training programme, the endosonographer examined 100 patients between July 1991 and April 1992, and another 131 patients between April 1992 and March 1993. Seventy-one tumours were resected (36 first period, 35 second) and therefore supplied the gold standard for calculating staging accuracy. The overall T-stage accuracy increased from 58% (21/36) in the first period to 83% (29/35) in the second. The overstaging rate did not vary between the two periods; however, fewer patients were understaged in the second period.

The majority of the tumours were T3 (45/71), with small numbers for the other three stages (T1 = 12, T2 = 8, T4 = 6). Hence no analysis of the variation between individual stages was appropriate. In addition, information on the division of the individual stages between the two periods was presented for T3 and T4 tumours only (T3: 21 first, 24 second; T4: 4 first, 2 second). Because T2 is suspected to be more difficult to stage, any imbalance between the distribution of these tumours between the two periods may have influenced the results.

TABLE 4.9 *Staging of metastases from gastro-oesophageal tumour studies*

Author and year of publication [reference]	No. patients and site of tumour	Conclusions
Binmoeller *et al.* 1995[3]	O = 38	25% understaged 7% overstaged
Botet *et al.* 1991[4]	O = 50	75% understaged No overstaging
Botet *et al.* 1991[5]	St = 50	92% correctly staged
François *et al.* 1996[11]	C = 29	75% understaged No overstaging
Tio *et al.* 1989[30]	St = 72	33% understaged No overstaging

TABLE 4.10 *Gastro-oesophageal grouped TNM staging studies*

Author and year of publication [reference]	No. patients and site of tumour	Stages included	Conclusions
Altorki *et al.* 1996[2]	C = 55	I; IIA; IIB; III	34 correct (62%) Stage II poorest
Botet *et al.* 1991[4]	O = 50	I; II; III; IV	30 correct (60%)
Botet *et al.* 1991[5]	St = 50	I; II; III; IV	39 correct (78%)
François *et al.* 1996[11]	C = 29	I; II; III; IV	21 correct (72%) Stage IV poorest
Tio *et al.* 1989[30]	St = 68	IA; IB; II; IIIA; IIIB; IV	41 correct (60%) Stage II poorest

4.2 EUS in gastro-oesophageal cancer – staging impact

4.2.1 Staging impact methodological studies

Only two studies were found that supplied any information on the staging impact of EUS, but both of these were designed to be more applicable to the analysis of therapeutic impact. These studies, by Jafri *et al.*[38] and Nickl *et al.*[39], are described in section 4.3 on therapeutic impact.

4.2.2 Staging impact comparative studies

Comparative studies were found only for incremental CT, not for comparison with other modalities. Eight of the studies addressing preoperative staging also supplied comparative performance results for CT. In all these studies both EUS and CT were performed on either the total study group or a smaller subgroup; results were compared with the single gold standard. Five studies assessed oesophageal cancer *(Table 4.11),* and three assessed gastric cancer *(Table 4.12)*. Of these, only five noted the correct limitation of CT (i.e. not being able to differentiate between tumour Stages T1 and T2) and hence combined T1 and T2. The authors of the other three studies, Ziegler *et al.* for oesophageal[35] and gastric[36] cancer, and Perng *et al.*[25] for gastric cancer, chose to combine tumour stages T2 and T3 and therefore cannot be directly compared.

As well as the small number of studies available, another factor that limited the amount of reliable quantitative data derived from direct comparisons of EUS and CT was the inadequate reporting of the comparative technology and the methodology. Of the five studies that, feasibly, could be compared, three did not supply information on the contrast protocol used and one also did not supply information on slice thickness. In all studies, only three pieces of information regarding CT were given, as listed in *Table 4.11* (i.e. CT scanner, slice thickness and contrast protocol). Regarding methodology, only three studies reported blinding between EUS and CT results (two of which were by Botet *et al.*[4,5]) and only the two studies by Botet *et al.*[4,5] analysed the complementary role of EUS and CT.

4.3 EUS in any clinical application – therapeutic impact

Four studies passed the preliminary inclusion criteria set for therapeutic impact. Due to their wide range of clinical applications, no data synthesis of any kind was performed. Descriptive summaries of the articles follow:

[37] Fok M, Cheng SWK, Wong J. Endosonography in patient selection for surgical treatment of esophageal carcinoma. *World J Surg* 1992;**16**:1098–103.

Consecutive and unselected patients with oesophageal cancer were staged by conventional methods. They were staged separately by EUS and then the two methods were compared for sensitivity, specificity and accuracy in assessing oesophageal involvement, oesophageal infiltration and lymph node metastases. The relevance of EUS staging to surgical management was analysed.

TABLE 4.11 *Staging impact comparative studies for oesophageal cancer*

	Botet *et al.* 1991[4]	**Greenberg *et al.* 1994**[13]	**Heintz *et al.* 1991**[17]	**Manzoni 1993**[20]	**Ziegler *et al.* 1991**[35]
No. receiving CT	42	20	22	54	37
No. receiving EUS	50	20	22	24	37
Demographics given for both	✗	✔	✔	✗	✔
CT scanner	GE 9800	GE 9800	DRH Siemens	DRH Siemens	Somatom DRG or DRH
Slice thickness	10 mm	10 mm	8 mm	N/S	8–10 mm
Contrast used	150–200 ml 60% iodinated 1.0 ml/s for 60 s 0.7 ml/s for rest	Intravenous and oral	Intravenous and oral	N/S	150–250 ml of Ultraquist® (UK: Ultravist®) 300 100–200 ml of 1–2% Gastrografin®
Stages combined	T1 and T2	No T1 patients	T1 and T2	T1 and T2	T2 and T3
Lymph node criteria	> 10 mm	N/S	> 10 mm	N/A	> 10 mm
Tumour	✔	✔	✔	✔	✔
Lymph nodes	✔	✔	✔	✗	✔
Metastases	✔	✗	✗	✗	✗
Overall stage	✔	✗	✗	✗	✗
Conclusions	CT most effective in staging M Combining EUS and CT most effective for overall stage	CT tends to understage tumours	EUS and CT comparable for T3/T4	EUS better than CT for preoperative T staging	EUS better than CT for preoperative TN staging

Patients were staged by both methods and underwent transthoracic resection. The overall accuracy of preoperative staging was significantly greater by EUS than by conventional methods ($p < 0.001$). EUS identified more advanced disease in 14 patients, as confirmed at surgery; however the choice of surgical management was not changed by EUS findings. In advanced disease, where surgery is used for curative treatment and palliation, EUS does not influence patient management.

[38] Jafri IH, Saltzman JR, Colby JM, Krims PE. Evaluation of the clinical impact of endoscopic ultrasonography in gastrointestinal disease. *Gastrointest Endosc* 1996;**44**:367–70.

In this prospective study clinicians completed a questionnaire both before requesting EUS and after the result was obtained. The questionnaires were compared to assess the impact of EUS in terms of changes in diagnosis, diagnostic certainty, management plans and whether EUS led to a more or less invasive course of therapy. A usefulness score for EUS was also awarded by the clinician. Diagnostic certainty and usefulness were graded from 1 to 5, with the highest score indicating the greatest clinical impact. All adults referred for EUS were entered into the study. Results were available for 63/67 patients. In 28, a change in treatment plan was attributable to EUS, the diagnostic certainty score increased from 2.8 to 4.3 after EUS ($p < 0.0001$), and the usefulness score for EUS was 4.1.

[39] Nickl NJ, Bhutani MS, Catalano M, Hoffman B, Hawes R, Chak A, *et al.* Clinical implications of endoscopic ultrasound: the American Endosonography Club Study. *Gastrointest Endosc* 1996;**44**:371–7.

In a similar study to Jafri *et al.*,[38] a prospective multicentre trial of 423 patients looked at changes in management plans and complication rates due to

TABLE 4.12 *Staging impact comparative studies for gastric cancer*

	Botet *et al.* 1991[5]	**Perng *et al.* 1996**[25]	**Ziegler *et al.* 1993**[36]
No. receiving CT	33	69	108
No. receiving EUS	50	69	108
Demographics given for both	✗	✔	✔
CT scanner	1200 SX GE 9800	Somatom DRH Siemens	Somatom DRG or DRH
Slice thickness	10 mm	8 mm	2 mm
Contrast used	150–200 ml of 60% iodinated 1.0 ml/s for 60 s 0.7 ml/s for rest 200 ml of Gastrografin	100–150 ml of Angiografin® 500 ml of 1–2% Gastrografin	150–250 ml of Ultraquist (UK: Ultravist) 300 100–200 ml of 1–2% Gastrografin
Stages combined	T1 and T2	T2 and T3	T2 and T3
Lymph node criteria	> 10 mm	> 10 mm	> 8 mm
Tumour	✔	✔	✔
Lymph nodes	✔	✔	✔
Metastases	✔	✗	✗
Overall stage	✔	✗	✗
Conclusions	EUS and CT were comparable in staging M Combining EUS and CT was more effective for overall stage than CT alone	EUS is the most reliable test for TN staging	EUS is the most reliable test for TN staging

EUS. All consecutive examinations were included. The results showed that the complication rate was 1.7%; it was 0.3% for severe complications. Minor symptoms after EUS were experienced by 31%.

In 74% of these patients, the management plans were altered by changing either the testing plan or the treatment plan. Major changes (e.g. surgery versus not surgery) occurred in 31%. Management changes were less expensive, invasive and risky in 55%; these factors were increased in 37% and remained the same in 8%. The impact of EUS was also analysed with regard to the clinical indication for EUS referral. Management changes were greatest in the evaluation of submucosal tumours and lowest in known anorectal cancer.

[42] Snady H, Cooperman A, Siegel J. Endoscopic ultrasonography compared with computed tomography with ERCP in patients with obstructive jaundice or small peripancreatic mass. *Gastrointest Endosc* 1992;**38**:27–34.

CT, ERCP and EUS were performed in 60 diagnostically problematic patients with obstructive jaundice or a peripancreatic mass. The EUS was performed with knowledge of the CT and ERCP results. This study compares the imaging modalities for detection of an abnormality, diagnosis and prediction of resectability. This was verified at surgery/biopsy in 42 patients and by clinical follow-up in 18.

EUS detected six false-positives and 13 false-negatives, resulting in changes in patient management. On the basis of the EUS findings, the decision was made to resect in two patients, not to resect in ten and to identify benign disease in seven. In 75%, EUS contributed to patient management by providing more details about the disease. It altered management in 32% by providing or changing the diagnosis.

4.4 EUS in any clinical application – patient outcome

Descriptive summaries of the two studies supplying information on patient outcome follow:

[40] Ramirez JM, Mortensen NJM, Takeuchi N, Smilgin Humphreys MM. Endoluminal ultrasonography in the follow-up of patients with rectal cancer. *Br J Surg* 1994;**81**:692–4.

In a study of 66 patients with a mean age of 68 years (range 43–87) who had undergone radical surgery for mid- and lower-third rectal carcinoma, EUS was used as a follow-up investigation. The results showed that EUS detected recurrence in 13 (20%) patients, three of which failed to be detected on digital examination and rigid sigmoidoscopy. Four patients were treated with salvage surgery and the remainder with palliative therapy. Six were alive at follow-up (4–50 months); four remained disease free, three of whom had undergone salvage surgery. A table is included setting out the clinical data of the patients with local recurrence.

EUS influenced management in three patients who underwent salvage surgery. It can detect local recurrence at an early stage and may therefore influence outcome in terms of survival.

[41] Setti Carraro P, Kamm MA, Nicholls RJ. Long-term results of postanal repair for neurogenic faecal incontinence. *Br J Surg* 1994;**81**:140–4.

This study looked at the long-term follow-up of 34 of 54 patients with neurogenic incontinence. Their median age at presentation was 64 years (range 30–83). A breakdown of clinical features is given in a table, subdivided by postoperative continence category. Patients underwent follow-up by clinical and anorectal physiological assessment at least 5 years after surgery. Anal EUS was performed in 30 patients, showing clinically undetected sphincter defects in 19. These results were unrelated to patient outcome measures of postoperative incontinence or postoperative anal pressures.

4.5 EUS in any clinical application – health economics

The two studies containing economic information are described next.

[43] Allgayer H. Cost-effectiveness of endoscopic ultrasonography in sub-mucosal tumours. *Gastrointest Endosc Clin North Am* 1991;**5**:625–9.

This was a retrospective study using data from 30 patients who were referred for further investigation of suspected submucosal tumours that had been identified on previous endoscopy. The comparator was CT scanning. The costs were restricted to investigation costs only, which were measured by hospital charges; the effectiveness measure was the number of correctly diagnosed patients. The study claimed that EUS is cost-effective. This may be true but the analysis presented used incorrect formulae and did not present marginal cost-effectiveness ratios. Some of the limitations of the study were recognised in the discussion.

[44] Prat F, Amouyal G, Amouyal P, Pelletier G, Fritsch J, Choury AD, *et al.* Prospective controlled study of endoscopic ultrasound and endoscopic retrograde cholangiography in patients with suspected common-bileduct lithiasis. *Lancet* 1996;**347**:75–9.

The aim of this study was to determine whether the use of EUS rather than endoscopic retrograde cholangiography (ERC) prevented unnecessary sphincterotomy or other surgery. A total of 119 patients underwent EUS and ERC carried out by independent investigators within 2 hours of each other, followed by endoscopic sphincterotomy as the gold standard for the presence or absence of gall-stones. A cost-analysis was presented, comparing three investigation strategies using resource data from the study, unit cost data from the literature, and charge data from French private clinics. Costs for investigations, hospitalisations and complications are included. The 'EUS first' strategy is more costly but, although the clinical data indicated that it may be more effective, no formal cost-effectiveness analysis was carried out.

Conclusion

In summary, in this chapter the results of the search strategy were presented, highlighting the retrieval performance of the search from the various resources available. Next, details of the studies included in the review were presented for each level of the Fineberg framework. For the staging impact levels these details were presented in tables. At the higher levels of the framework, where the studies were more varied in design, descriptive summaries were included. In the next chapter, tables are presented showing the primary reason why studies satisfying the initial inclusion criteria were excluded from description in this chapter and in the review. The results of the studies described in this chapter are given in chapter 6, where the results are also combined wherever possible.

Chapter 5

Details of studies excluded from the review

In this chapter the studies that satisfied the initial exclusion criteria, but which were excluded after reading the full article, are listed. The references for all studies listed in this chapter appear in the third section of the reference list.

5.1 EUS in gastro-oesophageal cancer – staging performance

Table 5.1 details the primary criteria that were used to exclude each of the studies that were not applicable to the review of preoperative staging accuracy of EUS for gastro-oesophageal cancer.[1–83] Only one criterion is indicated for each study; in many cases more than one may have been applicable but this was not further investigated. This table also includes those studies that were excluded from the subcategories of miniprobes and FNA.

5.2 EUS in gastro-oesophageal cancer – staging impact; 5.3 and 5.4 EUS in any clinical application – therapeutic impact and patient outcome

From the search for staging impact, therapeutic impact and patient outcome studies, 16 did not fulfil the inclusion criteria.[84–99] They are listed in *Table 5.2*.

5.5 EUS in any clinical application – health economics

The reason for excluding studies from the economics review was in all cases due to the absence of cost information. The references of the 30 studies excluded are given in the third section of the reference list.[100–129]

TABLE 5.1 *The 83 excluded staging performance studies*

Author and year (reference)	≤ 10 patients	No pathology gold standard	Insufficient raw data presented	No study information	Duplicate patient data
Aibe *et al.* 1986[1]	✔				
Aibe *et al.* 1992[2]			✔		
Akahoshi *et al.* 1992[3]			✔		
Akahoshi *et al.* 1995[4]	✔				
Andriulli *et al.* 1990[5]	✔				
Asaki *et al.* 1989[6]			✔		
Bandoh *et al.* 1993[7]			✔		
Binmoeller *et al.* 1994[8]	✔				
Boku *et al.* 1996[9]			✔		
Bolondi *et al.* 1987[10]			✔		
Catalano *et al.* 1994[11]			✔		
Catalano *et al.* 1995[12]			✔		
Chak *et al.* 1995[13]			✔		
Chandawarkar *et al.* 1996[14]			✔		
Chandawarkar *et al.* 1996[15]			✔		
Chonan *et al.* 1995[16]			✔		

continued

TABLE 5.1 contd *The 83 excluded staging performance studies*

Author and year (reference)	≤ 10 patients	No pathology gold standard	Insufficient raw data presented	No study information	Duplicate patient data
Dancygier and Classen 1986[17]			✔		
Dancygier and Classen 1989[18]	✔				
Date *et al.* 1990[19]			✔		
Fockens *et al.* 1994[20]	✔				
Francioni *et al.* 1987[21]			✔		
Frank *et al.* 1994[22]	✔				
Frank *et al.* 1996[23]	✔				
Fujishima *et al.* 1991[24]			✔		
Fujishima *et al.* 1996[25]	✔				
Fukuda 1986[26]				✔	
Glover *et al.* 1994[27]			✔		
Granstrom *et al.* 1993[28]			✔		
Grimm *et al.* 1992[29]	✔				
Grimm *et al.* 1992[30]			✔		
Grimm 1994[31]		✔			
Heyder and Lux 1986[32]			✔		
Hoffman *et al.* 1995[33]			✔		
Holden *et al.* 1996[34]			✔		
Kallimanis *et al.* 1995[35]		✔			
Maruta *et al.* 1994[36]			✔		
McLoughlin *et al.* 1995[37]	✔				
Melzer *et al.* 1995[38]	✔				
Meyer *et al.* 1994[39]				✔	
Mortensen *et al.* 1994[40]			✔		
Mortensen *et al.* 1995[41]			✔		
Mortensen *et al.* 1996[42]			✔		
Murata *et al.* 1987[43]					✔
Murata *et al.* 1996[44]			✔		
Murata *et al.* 1996[45]			✔		
Natsugoe *et al.* 1996[46]			✔		
Natsugoe *et al.* 1996[47]	✔				
Nousbaum *et al.* 1992[48]		✔			
Odegaard *et al.* 1990[49]	✔				
Ohashi *et al.* 1989[50]			✔		

continued

TABLE 5.1 contd *The 83 excluded staging performance studies*

Author and year (reference)	≤ 10 patients	No pathology gold standard	Insufficient raw data presented	No study information	Duplicate patient data
Okai *et al.* 1991[51]			✔		
Pedersen *et al.* 1996[52]	✔				
Rice *et al.* 1991[53]			✔		
Rosch and Classen 1990[54]		✔			
Rosch *et al.* 1992[55]			✔		
Saisho *et al.* 1995[56]			✔		
Shorvon *et al.* 1987[57]	✔				
Siewert *et al.* 1990[58]				✔	
Silva *et al.* 1988[59]			✔		
Smith *et al.* 1993[60]					✔
Strohm and Classen 1987[61]			✔		
Sugimachi *et al.* 1990[62]					✔
Takemoto *et al.* 1992[63]	✔				
Tanaka *et al.* 1990[64]	✔				
Tio and Tytgat 1984[65]	✔				
Tio *et al.* 1986[66]			✔		
Tio *et al.* 1989[67]					✔
Tio *et al.* 1989[68]					✔
Tio *et al.* 1990[69]					✔
Tio *et al.* 1990[70]			✔		
Tio *et al.* 1990[71]	✔				
Toh *et al.* 1993[72]			✔		
Van Dam *et al.* 1993[73]			✔		
Vilgrain *et al.* 1990[74]			✔		
Vilmann *et al.* 1991[75]			✔		
Wegener *et al.* 1994[76]	✔				
Wiersema *et al.* 1992[77]					✔
Yanai *et al.* 1993[78]					✔
Yanai *et al.* 1996[79]					✔
Yasuda *et al.* 1986[80]	✔				
Yasuda *et al.* 1992[81]			✔		
Yasuda *et al.* 1996[82]			✔		
Yoshikane *et al.* 1994[83]			✔		
Total	**22**	**4**	**45**	**3**	**9**

TABLE 5.2 *The 16 excluded staging impact, therapeutic impact and patient outcome studies*

Author and year (reference)	No original data	Not a comparative study	No link between EUS and decision/outcome
Beynon *et al.* 1989[84]	✔		
Cahn *et al.* 1996[85]			✔
Chak *et al.* 1995[86]		✔	
Dill *et al.* 1995[87]			✔
Felt-Bersma *et al.* 1996[88]			✔
Herzog *et al.* 1994[89]		✔	
Meyenberger *et al.* 1996[90]			✔
Mosnier *et al.* 1990[91]			✔
Motoo *et al.* 1995[92]		✔	
Nakao *et al.* 1994[93]		✔	
Nickl and Cotton 1990[94]			✔
Nielsen *et al.* 1993[95]			✔
Nielsen *et al.* 1994[96]			✔
Solomon *et al.* 1994[97]		✔	
Taal *et al.* 1989[98]		✔	
Tio *et al.* 1994[99]		✔	
Total	**1**	**7**	**8**

Conclusion

In this chapter the studies that were excluded from the review have been listed. In the next chapter we return to the included studies and consider their findings.

Chapter 6
Results

The results for the staging performance of EUS in gastro-oesophageal cancer are presented in sections 6.1–6.4.

The results for the comparative CT staging impact studies are presented in section 6.5. No results are presented in this chapter for the higher levels of the Fineberg framework. The methodology used to combine results was described in section 3.1.5.

6.1 Tumour staging

Of 27 independent studies identified, three gave sufficient information to separate the results into the two categories of oesophagus and stomach. Hence a total of 30 results were available: 13 oesophageal; 13 gastric; and four cardia or gastro-oesophageal junction. The quantitative analysis was performed in three sections: studies with results for oesophageal cancer staging; studies with results for gastric cancer staging; and, finally, all studies combined together, with an analysis of the influence of the particular anatomical site on the results. Owing to the small number of studies addressing cancer specific to the cardia or gastro-oesophageal junction, no separate analysis of the accuracy of tumour staging for this site was performed.

6.1.1 Oesophageal tumour staging

Thirteen studies supplied independent results for the accuracy of staging oesophageal tumour infiltration. *Table 6.1* presents the raw data extracted from the studies along with the statistical results. FPR2, TPR2 and OR2 represent the resulting statistics from the addition of 0.5 to the true-positive (TP), false-negative (FN), false-positive (FP) and true-negative (TN), as recommended by Irwig *et al.*[26] and Moses *et al.*[25]

For the three stages of the analysis described in section 3.1.5, the results are shown in *Figures 6.1–6.3*. As no points were outside the clinically significant range (TPR > 0.5; FPR < 0.5), all were included in the analysis. The gradient

TABLE 6.1 *Statistical results of oesophageal tumour staging EUS studies*

Author and year of publication [reference]	TP	FN	FP	TN	FPR	TPR	FPR2	TPR2	OR	OR2
Binmoeller *et al.* 1995[3]	8	2	1	27	0.036	0.800	0.052	0.773	108.000	62.333
Botet *et al.* 1991[4]	5	2	0	43	0.000	0.714	0.011	0.688	–	191.400
Catalano *et al.* 1994[7]	30	4	7	59	0.106	0.882	0.112	0.871	63.214	53.770
Dittler *et al.* 1993[9]	54	4	7	102	0.064	0.931	0.068	0.924	196.714	165.519
Grimm *et al.* 1993[14]	23	2	1	37	0.026	0.920	0.038	0.904	425.500	235.000
Heintz *et al.* 1991[17]	8	1	2	11	0.154	0.889	0.179	0.850	44.000	26.067
Hünerbein *et al.* 1996[19]	4	1	0	14	0.000	0.800	0.033	0.750	–	87.000
Manzoni 1993[20]	7	2	0	15	0.000	0.778	0.031	0.750	–	93.000
Murata *et al.* 1988[22]	82	5	2	83	0.024	0.943	0.029	0.938	680.600	501.000
Murata *et al.* 1993[23]	152	14	8	143	0.053	0.916	0.056	0.913	194.071	177.556
Peters *et al.* 1994[26]	9	2	5	18	0.217	0.818	0.229	0.792	16.200	12.782
Takemoto *et al.* 1986[29]	9	0	1	2	0.333	1.000	0.375	0.950	–	31.667
Ziegler *et al.* 1991[35]	8	1	1	27	0.036	0.889	0.052	0.850	216.000	103.889

–, not evaluable when either TPR or FPR = 1

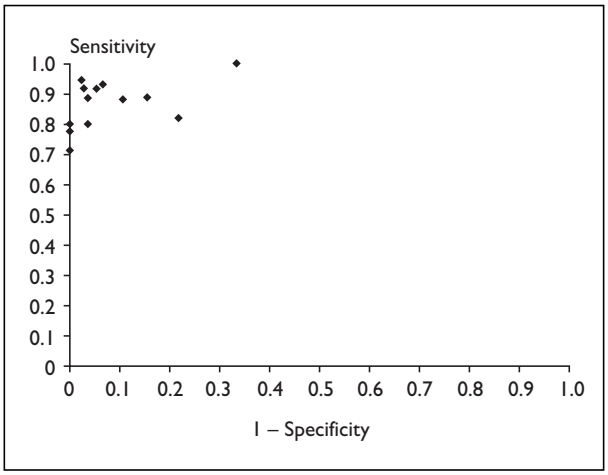


FIGURE 6.1 *ROC scatterplot – oesophageal tumour staging*

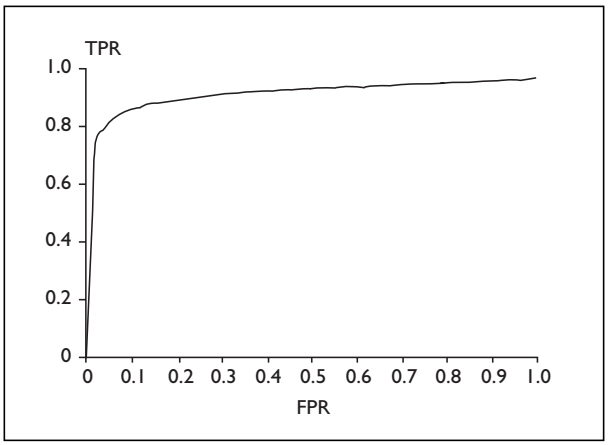


FIGURE 6.3 *EWLS ROC curve – oesophageal tumour staging*

and intercept for the EWLS and RR fits are shown in *Table 6.2,* together with the summary estimate Q*. The standard error (Se) for the EWLS method is quoted but it was not possible to calculate an equivalent error for the RR method, so only the EWLS method is used further in the analysis.

6.1.2 Gastric tumour staging

Thirteen studies supplied results for the accuracy of staging gastric tumour infiltration. The raw data extracted from the studies together with the corresponding calculated statistics are shown in *Table 6.3.*

No points fell outside the specified range on the ROC scatterplot *(Figure 6.4),* hence all were included in the analysis *(Figures 6.5* and *6.6). Table 6.4* illustrates the results.

6.1.3 Gastro-oesophageal tumour staging

An overall analysis was performed by combining all the results of the three anatomical locations: oesophagus, stomach, and the cardia or

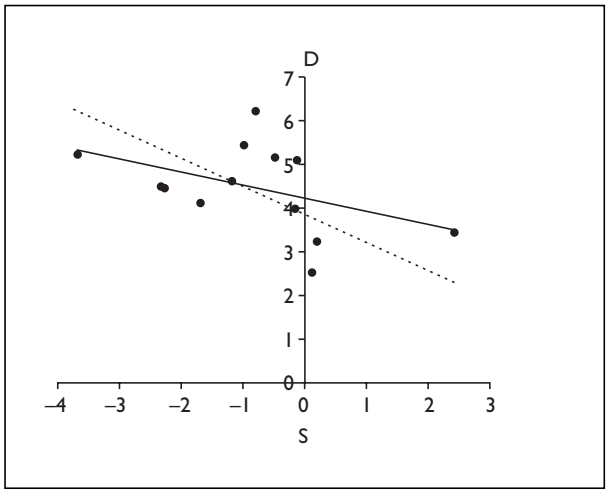


FIGURE 6.2 *SROC curve – oesophageal tumour staging (●, D; ——, EWLS total; - - -, RR total)*

TABLE 6.2 *Oesophageal tumour staging EUS results studies*

Method	B	A	Se(A)	Q*
RR total	–0.65	3.88	–	0.87
EWLS total	–0.30	4.23	0.30	0.89
See section 3.1.5. for method				

gastro-oesophageal junction. The combination of these three sites will be referred to as gastro-oesophageal. Twenty-seven studies supplied 30 sets of results for the accuracy of staging gastro-oesophageal tumour infiltration. The raw data extracted from the studies together with the corresponding calculated statistics are shown in *Tables 6.1* and *6.3* for the oesophagus and stomach respectively. *Table 6.5* presents only the additional results for the remaining studies addressing the accuracy of staging cardia or gastro-oesophageal junction cancers.

From *Figure 6.7* it can be seen that only one point was outside the range (TPR > 0.5; FPR < 0.5). This was the study by Hordijk *et al.*[18] In order to assess the influence of excluding outlying studies, as suggested by Moses *et al.*,[25] compared with including the complete data set, as suggested by Irwig *et al.*,[26] both methods were used to plot an SROC curve for the data *(Figure 6.8).* The results of fitting lines to the points of the SROC curve are shown in *Table 6.6.* For the remainder of the analysis, the EWLS method was used on the complete data set and Figure 6A shows the corresponding ROC curve.

Table 6.7 provides an overall summary of the Q* values for each of the three topics for both the

TABLE 6.3 *Statistical results of gastric tumour staging studies*

Author and year of publication [reference]	TP	FN	FP	TN	FPR	TPR	FPR2	TPR2	OR	OR2
Akahoshi *et al.* 1991[1]	59	2	0	13	0.000	0.967	0.036	0.960	–	642.600
Botet *et al.* 1991[5]	11	1	1	37	0.026	0.917	0.038	0.885	407.000	191.667
Caletti *et al.* 1993[6]	10	1	2	22	0.083	0.909	0.100	0.875	110.000	63.000
Dittler *et al.* 1993[10]	65	14	11	164	0.063	0.823	0.065	0.819	69.221	64.616
Grimm *et al.* 1993[14]	80	14	3	50	0.057	0.851	0.065	0.847	95.238	80.103
Hünerbein *et al.* 1996[19]	19	9	4	28	0.125	0.679	0.136	0.672	14.778	13.000
Massari *et al.* 1996[21]	26	0	2	37	0.051	1.000	0.063	0.981	–	795.000
Murata *et al.* 1988[22]	100	5	3	38	0.073	0.952	0.083	0.948	253.333	201.000
Perng *et al.* 1996[25]	26	4	2	35	0.054	0.867	0.066	0.855	113.750	83.622
Saito *et al.* 1991[27]	56	4	1	49	0.020	0.933	0.029	0.926	686.000	414.333
Shimizu *et al.* 1994[28]	84	6	2	36	0.053	0.933	0.064	0.929	252.000	189.800
Tio *et al.* 1989[30]	27	1	3	37	0.075	0.964	0.085	0.948	333.000	196.429
Ziegler *et al.* 1993[36]	50	4	4	50	0.074	0.926	0.082	0.918	156.250	125.938
–, not evaluable when either TPR or FPR = 1										

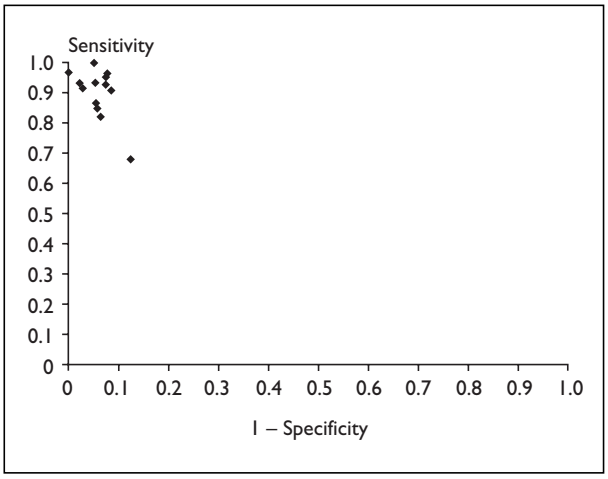


FIGURE 6.4 *ROC scatterplot – gastric tumour staging*

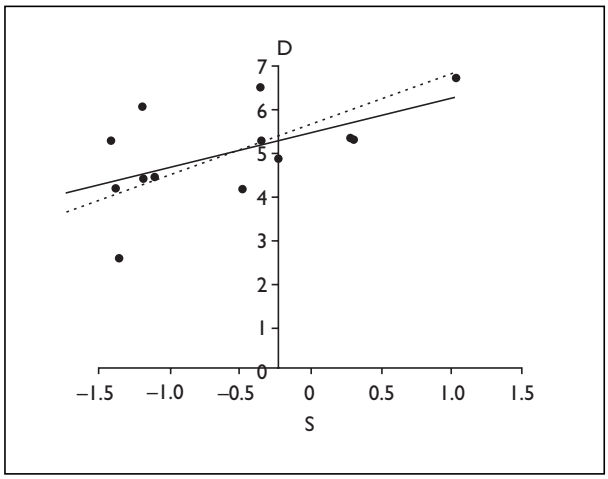


FIGURE 6.5 *SROC curve – gastric tumour staging (●, D; ——, EWLS total; - - -, RR total)*

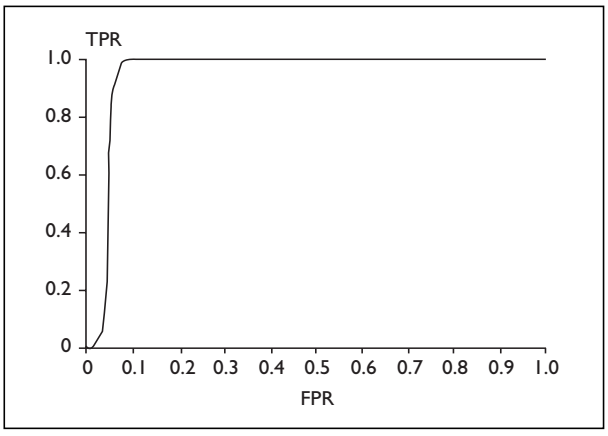


FIGURE 6.6 *EWLS ROC curve – gastric tumour staging*

TABLE 6.4 *Gastric tumour staging EUS results*

Method	B	A	Se(A)	Q*
RR total	1.16	5.38	–	0.94
EWLS total	0.80	5.26	0.29	0.93
See section 3.1.5. for method				

TABLE 6.5 *Statistical results of cardia or gastro-oesophageal junction tumour staging*

Author and year of publication [reference]	TP	FN	FP	TN	FPR	TPR	FPR2	TPR2	OR	OR2
Altorki *et al.* 1996[2]	15	2	8	28	0.222	0.882	0.230	0.861	26.250	20.788
François *et al.* 1996[11]	11	1	2	15	0.118	0.917	0.139	0.885	82.500	47.533
Greenberg *et al.* 1994[13]	5	1	1	13	0.071	0.833	0.100	0.786	65.000	33.000
Hordijk *et al.* 1993 [18]	5	7	0	27	0.000	0.417	0.018	0.423	–	40.333
–, not evaluable when either TPR or FPR = 1										

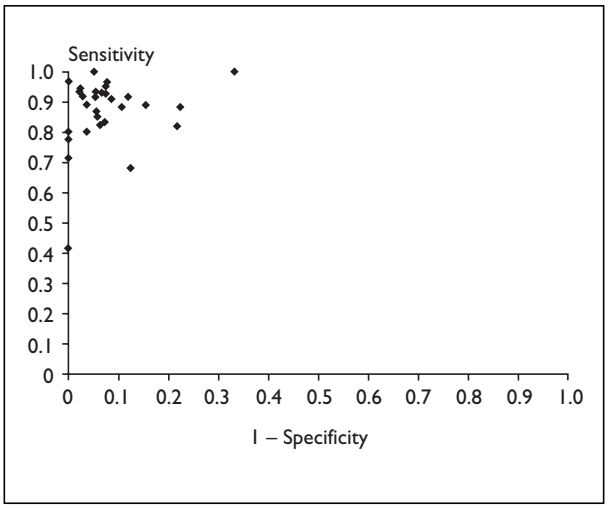


FIGURE 6.7 *ROC scatterplot – gastro-oesophageal tumour staging*

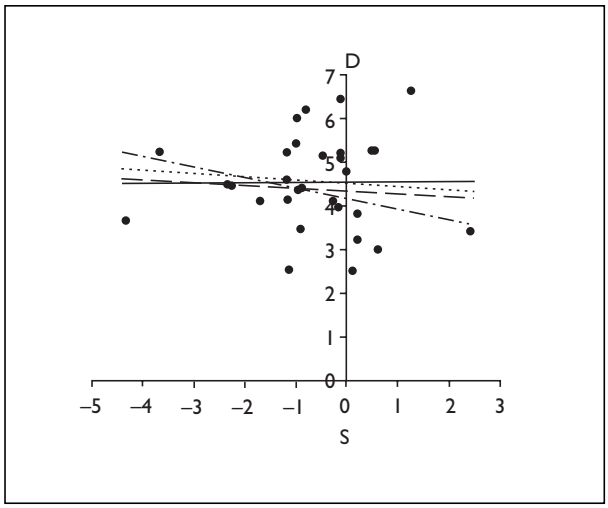


FIGURE 6.8 *SROC curve – gastro-oesophageal tumour staging (●, D; ——, EWLS total; – – –, RR total; - - - , EWLS excluded; — · —, RR excluded)*

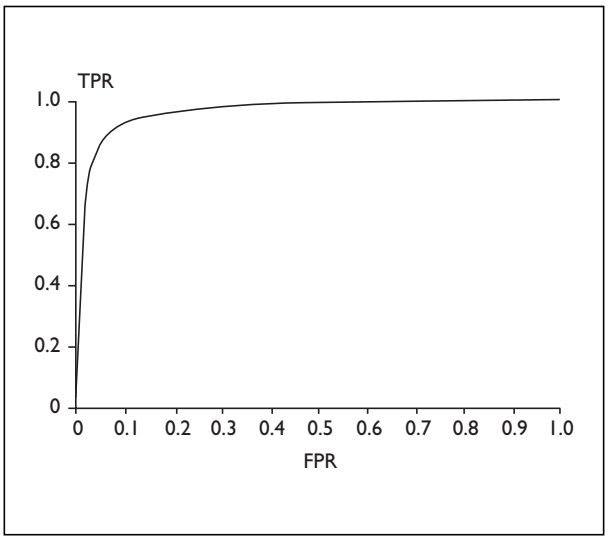


FIGURE 6.9 *EWLS ROC curve – gastro-oesophageal tumour staging for complete data set*

TABLE 6.6 *Gastro-oesophageal tumour staging EUS results*

Method	B	A	Se(A)	Q*
RR total	–0.06	4.38	–	0.90
EWLS total	0.01	4.58	0.22	0.91
RR excluded	–0.24	4.2	–	0.90
EWLS excluded	–0.08	4.56	0.23	0.91
See section 3.1.5. for method				

TABLE 6.7 *Summary of Q* values for each topic, including the complete data set*

Topic	Q* EWLS	Q* RR
Oesophageal tumour staging	0.89	0.87
Gastric tumour staging	0.93	0.94
Gastro-oesophageal tumour staging	0.91	0.90

TABLE 6.8 *Differentiation between subdivisions of T1 using miniprobes vs using conventional EUS probes*

Author and year of publication [reference]	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	OR
Differentiation between subdivisions of T1 – using miniprobes						
Hasegawa *et al.* 1996[16]	85.7	94.4	85.7	94.4	92.0	102
Murata *et al.* 1996[24]	88.0	100	100	87.0	93.3	–
Yanai *et al.* 1996[34]	75.0	87.5	96.4	43.8	77.3	21
Differentiation between subdivisions of T1 – using conventional EUS probes						
Hasegawa *et al.* 1996[16]	71.4	87.5	71.4	87.5	82.6	17.5

–, not evaluable when either TPR or FPR = 1

TABLE 6.9 *Differentiation between T1 and not T1 using miniprobes*

Author and year of publication [reference]	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	OR
Murata *et al.* 1996[24]	91.8	100	100	50.0	92.5	–

–, not evaluable when either TPR or FPR = 1

TABLE 6.10 *Statistical results of lymph node staging studies associated with primary oesophageal tumours*

Author and year of publication [reference]	TP	FN	FP	TN	FPR	TPR	FPR2	TPR2	OR	OR2
Binmoeller *et al.* 1995[3]	26	3	5	4	0.556	0.897	0.550	0.883	6.933	6.195
Botet *et al.* 1991[4]	35	1	5	9	0.357	0.972	0.367	0.959	63.000	40.879
Catalano *et al.* 1994[7]	57	7	2	34	0.056	0.891	0.068	0.885	138.429	105.800
Dittler *et al.* 1993[9]	85	29	16	37	0.302	0.746	0.306	0.743	6.778	6.587
Grimm *et al.* 1993[14]	37	3	5	17	0.227	0.925	0.239	0.915	41.933	34.091
Heintz *et al.* 1991[17]	13	2	0	4	0.000	0.867	0.100	0.844	–	48.600
Hünerbein *et al.* 1996[19]	13	1	1	2	0.333	0.929	0.375	0.900	26.000	15.000
Peters *et al.* 1994[26]	22	2	6	4	0.600	0.917	0.591	0.900	7.333	6.231
Ziegler *et al.* 1991[35]	16	9	3	9	0.250	0.640	0.269	0.635	5.333	4.714

–, not evaluable when either TPR or FPR = 1

EWLS and RR fits, using the complete data set in each case.

6.1.4 Tumour staging using miniprobes

The performance results for the differentiation of mucosal from submucosal cancer by miniprobes are shown in *Table 6.8*. In addition, one study compared miniprobes with conventional EUS probes for differentiating between the subdivisions of T1; these results are shown for comparison in *Table 6.8*. One study supplied information for the performance of miniprobes in differentiating T1 tumours from tumours classified as T2 or above (*Table 6.9*).

6.2 Lymph node staging

Of the 27 studies, 18 provided information on the performance of EUS for lymph node staging. The following three sections (6.2.1–6.2.3) supply the results for each of the three anatomical categories.

6.2.1 Lymph node staging of primary oesophageal tumours

Nine oesophageal studies also provided sufficient data to extract results of the accuracy of staging lymph nodes. The statistical results of these studies are given in *Table 6.10*.

By plotting the ROC scatterplot of the data it could be seen that there were two outliers with (1 – specificity) > 0.5 *(Figure 6.10)*. However, owing to the small total number of studies in this case it was decided not to exclude these outliers[25] and to perform the analysis on the complete data set. *Figure 6.11* shows the resulting SROC curve and *Figure 6.12* the ROC curve for the EWLS fit. *Table 6.11* contains the results of the fitted lines and Q*.

6.2.2 Lymph node staging of primary gastric tumours

Eight studies provided raw data for the extraction of results concerning the accuracy of staging lymph nodes. The statistical results of these studies are shown in *Table 6.12*.

On plotting the ROC scatterplot of the data it could be seen that there was one outlier with (1 – specificity) > 0.5 *(Figure 6.13)*. Owing to the small total number of studies it was decided not to exclude this outlier and to perform the analysis on the complete data set. The results are shown in *Figures 6.14* and *6.15*, and in *Table 6.13*.

6.2.3 Lymph node staging of primary gastro-oesophageal tumours

Twenty gastro-oesophageal studies also provided sufficient data to extract statistical results of the accuracy of staging lymph nodes. These are presented in *Tables 6.10* and *6.12* for the oesophagus and stomach respectively; *Table 6.14* shows the results for the cardia studies.

On the ROC scatterplot of the complete data set it could be seen that there were three outliers with (1 – specificity) > 0.5 *(Figure 6.16)*. In order to assess the influence of excluding outlying studies, as suggested by Moses *et al.*,[25] compared with including the complete data set, as suggested by Irwig *et al.*,[26] both methods were used to plot an SROC curve for the data *(Figure 6.17)*. The results of fitting lines to the

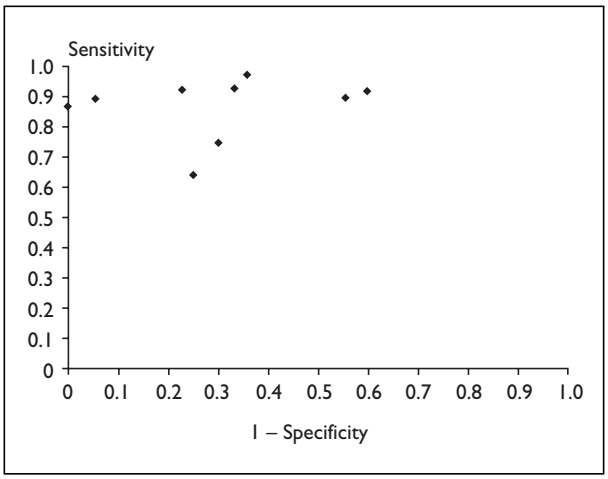


FIGURE 6.10 *ROC scatterplot – lymph node staging of primary oesophageal tumours*

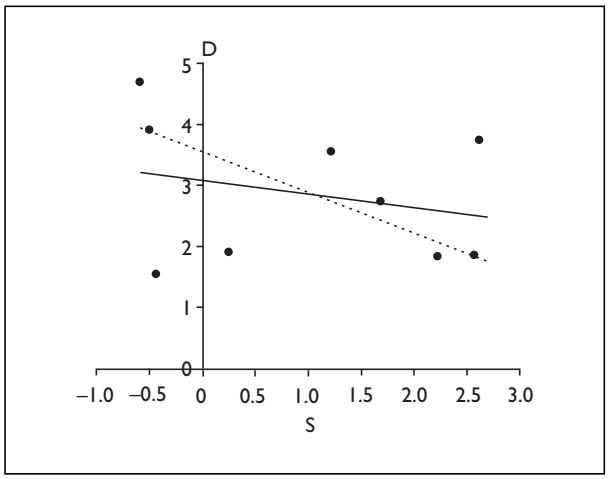


FIGURE 6.11 *SROC curve – lymph node staging of primary oesophageal tumours (●, D; ——, EWLS total; - - -, RR total)*

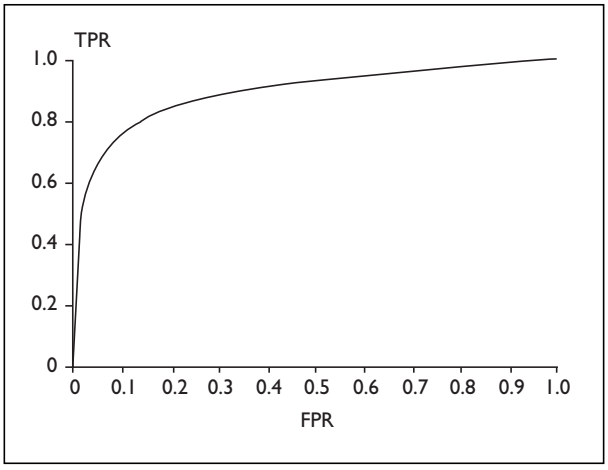


FIGURE 6.12 *EWLS ROC curve – lymph node staging of primary oesophageal tumours*

TABLE 6.11 *Lymph node staging EUS results for primary oesophageal tumour studies*

Method	B	A	Se(A)	Q*
RR total	–0.67	3.54	–	0.85
EWLS total	–0.22	3.07	0.50	0.82
See section 3.1.5. for method				

TABLE 6.12 *Statistical results of lymph node staging studies associated with primary gastric tumours*

Author and year of publication [reference]	TP	FN	FP	TN	FPR	TPR	FPR2	TPR2	OR	OR2
Botet *et al.* 1991[5]	31	8	1	10	0.091	0.795	0.125	0.788	38.750	25.941
Dittler *et al.* 1993[10]	130	53	5	66	0.070	0.710	0.076	0.709	32.377	29.493
Grimm *et al.* 1993[14]	60	13	9	49	0.155	0.822	0.161	0.818	25.128	23.351
Hünerbein *et al.* 1996[19]	24	10	5	15	0.250	0.706	0.262	0.700	7.200	6.576
Massari *et al.* 1996[21]	40	13	5	7	0.417	0.755	0.423	0.750	4.308	4.091
Perng *et al.* 1996[25]	25	12	8	24	0.250	0.676	0.258	0.671	6.250	5.880
Tio *et al.* 1989[30]	36	6	15	15	0.500	0.857	0.500	0.849	6.000	5.615
Ziegler *et al.* 1993[36]	40	18	6	44	0.120	0.690	0.127	0.686	16.296	14.988

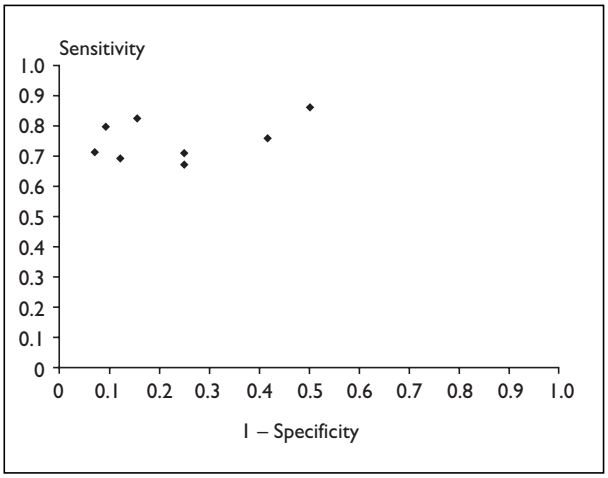


FIGURE 6.13 *ROC scatterplot – lymph node staging of primary gastric tumours*

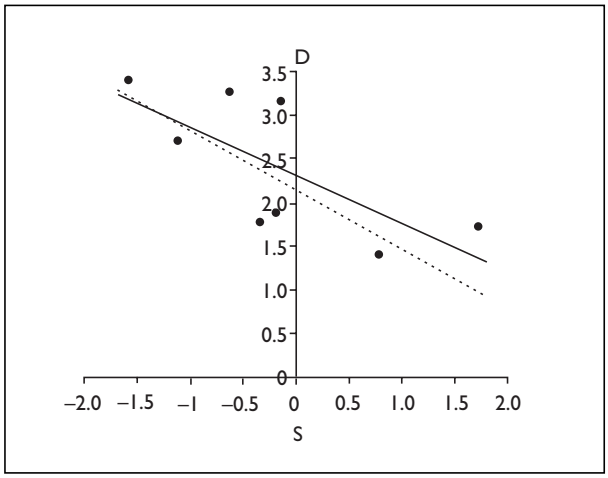


FIGURE 6.14 *SROC curve – lymph node staging of primary gastric tumours (●, D; ——, EWLS total; - - -, RR total)*

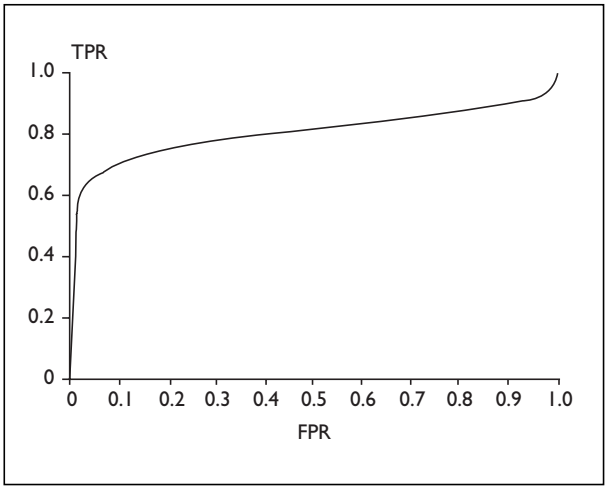


FIGURE 6.15 *EWLS ROC curve – lymph node staging of primary gastric tumours*

TABLE 6.13 *Lymph node staging EUS results for primary gastric tumour studies*

Method	B	A	Se(A)	Q*
RR total	–0.67	2.15	–	0.75
EWLS total	–0.54	2.31	0.22	0.76
See section 3.1.5. for method				

TABLE 6.14 *Statistical results of lymph node staging studies associated with primary tumours at the cardia*

Author and year of publication [reference]	TP	FN	FP	TN	FPR	TPR	FPR2	TPR2	OR	OR2
Altorki *et al.* 1996[2]	6	4	0	6	0.000	0.600	0.071	0.591	–	18.778
François *et al.* 1996[11]	16	3	0	10	0.000	0.842	0.045	0.825	–	99.000
Greenberg *et al.* 1994[13]	22	15	6	12	0.333	0.595	0.342	0.592	2.933	2.792
–, not evaluable when either TPR or FPR = 1										

points of the SROC curve are shown in *Table 6.15, and the ROC curve for the EWLS fit is given in Figure 6.18.* For the EWLS plots, it was decided to use both the full data set and the reduced data set to complete the analysis, allowing full comparison of the alternative methodologies. *Figure 6.18* shows the EWLS total data set, and *Figure 6.19* the EWLS for the data set with the three outlying results excluded. To compare the influence of the two methods of fitting lines with the SROC curve, the results from the robust RR method were also used to plot the ROC curve, for both the total data set *(Figure 6.20)* and the data set with the three outlying results excluded *(Figure 6.21).* The results are given in *Table 6.15.*

Table 6.16 provides an overall summary of the Q* values for each of the three topics for both the EWLS and RR fits, using the complete data set in each case.

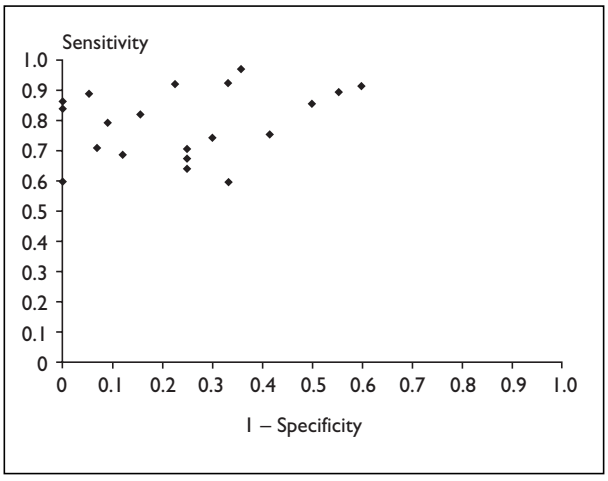


FIGURE 6.16 *ROC scatterplot – lymph node staging of primary gastro-oesophageal tumours*

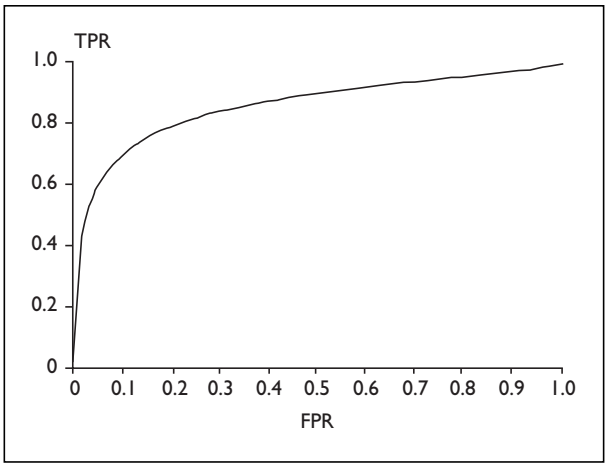


FIGURE 6.18 *EWLS ROC curve for total data set – lymph node staging of primary gastro-oesophageal tumours (all 20 studies included)*

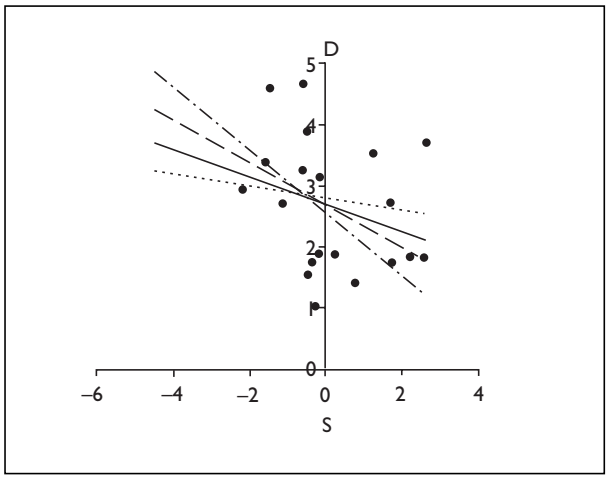


FIGURE 6.17 *SROC curve – lymph node staging of primary gastro-oesophageal tumours (●, D; ——, EWLS total; — —, RR total; - - - -, EWLS excluded; — · —, RR excluded)*

TABLE 6.15 *Lymph node staging EUS results for primary gastro-oesophageal tumour studies*

Method	B	A	Se(A)	Q*
RR total	–0.34	2.70	–	0.79
EWLS total	–0.22	2.71	0.24	0.79
RR excluded	–0.51	2.57	–	0.78
EWLS excluded	–0.10	2.81	0.28	0.80
See section 3.1.5. for method				

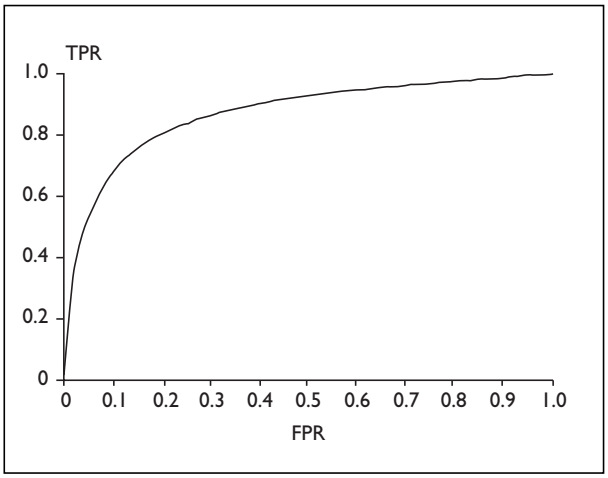


FIGURE 6.19 *EWLS ROC curve for data set – lymph node staging of primary gastro-oesophageal tumours (17 of 20 studies included)*

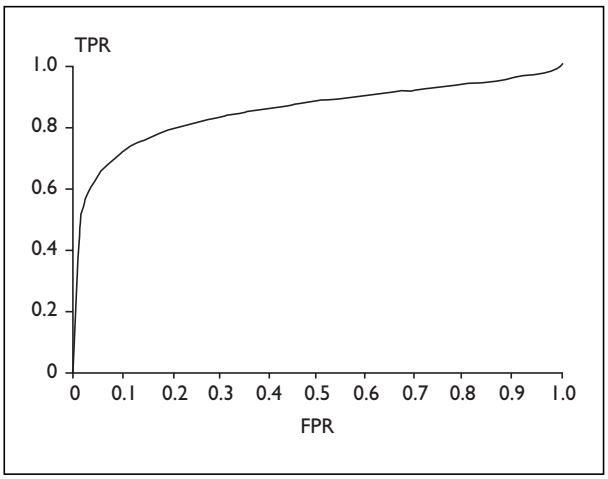


FIGURE 6.20 *RR ROC curve for total data set – lymph node staging of primary gastro-oesophageal tumours*

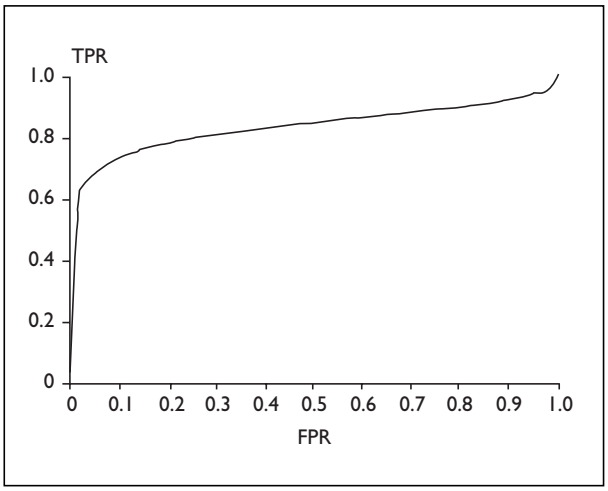


FIGURE 6.21 *RR ROC curve for total data set with outlying results excluded – lymph node staging of primary gastro-oesophageal tumours*

TABLE 6.16 *Summary of Q* values from the complete data set for each topic*

Topic	Q* EWLS	Q* RR
Lymph nodes associated with oesophageal tumours	0.82	0.85
Lymph nodes associated with gastric tumours	0.76	0.75
Lymph nodes associated with gastro-oesophageal tumours	0.79	0.79

6.3 Staging of metastases

The results of the five studies supplying information for the staging of metastases are shown in *Table 6.17;* the raw data are supplied in appendix 4. Combination of these results further into an overall summary estimate was not performed owing to the inherent clinical limitations of EUS for staging metastases, as discussed further in chapter 8.

6.4 Grouped TNM staging

Of the five studies identified that supplied information on grouped staging of the individual TNM stages, none supplied complete information on all stages. Therefore, no dichotomy was performed in order to calculate the summary statistics. The raw data are supplied in appendix 4.

6.5 Comparative CT staging performance

Eight of the 27 EUS studies (30%) also addressed the performance of CT.

6.5.1 CT tumour staging

Of the eight CT studies, five provided sufficient information to assess the performance of CT in differentiating T1 or T2 tumours from T3 or T4 tumours. No meta-analysis of these results, which are shown individually in *Table 6.18,* was performed. The raw data for these and other differentiations between the T stages are shown in appendix 4.

6.5.2 CT lymph node staging

Only seven studies reported the performance of CT for lymph node staging *(Table 6.19).*

TABLE 6.17 *Results of EUS staging of metastases*

Metastases	Binmoeller et *al.* 1995[3]	Botet et *al.* 1991[4]	Botet et *al.* 1991[5]	François et *al.* 1996[11]	Tio et *al.* 1989[30]
Sensitivity	75.0	25.0	N/A	25.0	66.7
Specificity	93.5	100.0	N/A	100.0	100.0
PPV	60.0	100.0	N/A	100.0	100.0
NPV	96.7	66.7	N/A	89.3	97.1
Accuracy	91.4	70.0	N/A	89.7	97.2
OR	43.5	N/A	N/A	N/A	N/A

N/A, not available

TABLE 6.18 *Performance statistics for CT tumour staging*

T1/T2	Botet et *al.* 1991[4]	Botet et *al.* 1991[5]	Greenberg et *al.* 1994[13]	Heintz et *al.* 1991[17]	Manzoni 1993[20]
Sensitivity	40.0	66.7	83.3	55.6	80.0
Specificity	97.1	66.7	14.3	84.6	86.2
PPV	66.7	33.3	29.4	71.4	83.3
NPV	91.9	88.9	66.7	73.3	83.3
Accuracy	90.0	66.7	35.0	72.7	83.3
OR	22.7	4.0	0.8	6.9	25.0

TABLE 6.19 *Performance statistics for CT lymph node staging*

Nodes	Botet et *al.* 1991[4]	Botet et *al.* 1991[5]	Greenberg et *al.* 1994[13]	Heintz et *al.* 1991[17]	Perng et *al.* 1996[25]	Ziegler et *al.* 1991[35]	Ziegler et *al.* 1993[36]
Sensitivity	79.3	67.9	50.0	66.7	44.4	40.0	44.8
Specificity	61.5	60.0	66.7	25.0	49.1	66.7	58.0
PPV	82.1	90.5	71.4	76.9	30.8	71.4	55.3
NPV	57.1	25.0	44.4	16.7	63.4	34.8	47.5
Accuracy	73.8	66.7	56.3	57.9	47.5	48.6	50.9
OR	6.1	3.2	2.0	0.7	0.8	1.3	1.1

6.5.3 CT staging of metastases

Both studies by Botet *et al.*[4,5] reported the results of CT staging of metastases, but only the oesophageal study supplied sufficient information to calculate the performance statistics *(Table 6.20)*. None of the other studies reported any information on CT metastases staging.

TABLE 6.20 *Performance statistics for CT staging of metastases*

Metastases	Botet et *al.* 1991[4]
Sensitivity	75.0
Specificity	100.0
PPV	100.0
NPV	86.7
Accuracy	90.5
OR	N/A

6.5.4 CT grouped TNM staging

The two studies by Botet *et al.* both reported the performance of CT for the grouped TNM staging, with no subdivision of stages into A and B *(Figure 3.2b)*. Uniquely, both Botet's studies investigated the

TABLE 6.21 *Grouped TNM staging of oesophageal cancer*

Oesophagus	**EUS + CT**			**CT**		
Botet *et al.* 1991[4]	**I or II**	**III**	**IV**	**I or II**	**III**	**IV**
Sensitivity	77.8	94.1	81.3	55.6	58.8	75.0
Specificity	97.0	80.0	100.0	81.8	72.0	100.0
PPV	87.5	76.2	100.0	45.5	58.8	100.0
NPV	94.1	95.2	89.7	87.1	72.0	86.7
Accuracy	92.9	85.7	92.9	76.2	66.7	90.5
OR	112.0	64.0	N/A	5.6	3.7	N/A

TABLE 6.22 *Grouped TNM staging of gastric cancer*

Stomach	**EUS + CT**			**CT**		
Botet *et al.* 1991[5]	**I or II**	**III**	**IV**	**I or II**	**III**	**IV**
Sensitivity	90.9	66.7	57.1	54.5	46.7	33.3
Specificity	77.3	77.8	100.0	77.3	64.7	81.5
PPV	66.7	71.4	100.0	54.5	53.8	28.6
NPV	94.4	73.7	89.7	77.3	57.9	84.6
Accuracy	81.8	72.7	90.9	69.7	56.3	72.7
OR	34.0	7.0	N/A	4.1	1.6	2.2

combined influence of EUS and CT for the grouped TNM staging, and these data were compared with the performance of CT alone. The performance results of this comparison are shown in *Table 6.21* for oesophageal cancer[4] and *Table 6.22* for gastric cancer.[5] The raw data are supplied in appendix 4.

Conclusion

In this chapter the results concerning staging performance reported by the studies included in the systematic review have been presented. Where possible, the results have been combined using the methodology outlined in section 3.1.5 to express the performance in terms of a single statistic (Q*). The limited results available relating to staging impact and the use of CT for staging were also presented. In the next chapter, further analysis is performed to determine whether significant differences between studies exist owing to methodological or study design factors.

Chapter 7

Analysis of the robustness of the results

If all the studies were exactly alike and free from differences due to factors related to their performance and threats to validity in the form of biases, then the results presented in chapter 6 would be reliable evidence of the true staging performance of EUS. Each ROC curve would provide the range of possible operating points for the test, with Q* being the optimum threshold to balance sensitivity and specificity. Throughout the review it has been emphasised that this is not the case.

The poor quality of the original studies, in terms of study design and, in particular, the completeness of the subsequent published reports, makes reliance on their results at face value controversial. It is necessary to attempt an individual assessment of their validity and homogeneity to allow the extraction of a more reliable estimate from their results. Using the checklists described in section 3.1.3, each study was assessed for the presence of risk of bias and the qualitative factors of the study were noted.

The methodology for determining if the results of the study were related to the likely risk of one or more biases and/or factors was described in section 3.1.6 and expands on the summary ROC curves already presented. Four biases and six factors were evaluated *(Table 7.1)*. Note that the blinding biases refer to the four biases categorised under the heading 'Independence of interpretations' in *Table 3.5*. The report of any blinding of any information for any of these four biases correlated with a negative response for the risk of bias in the regression analysis.

The full list of biases and factors could not be evaluated because some did not vary between studies. For example, all 30 studies supplied insufficient information adequately to assess the risk of referral bias, whereas all were free from incorporation bias. To apply the biases and factors to a multiple regression it was necessary to divide their values into two categories representing the binary outcomes 0 or 1; this is also shown in *Table 7.1*. All ten variables were placed in a multiple regression with a 5% ($p < 0.05$) level of significance for inclusion. No variables were found to be significant. Regression was then performed on each variable using the same level of significance ($p < 0.05$). Care must be exercised in selecting a level of

TABLE 7.1 *Bias risks and factors analysed, and how values were assigned to one of two categories; this division is described as the dichotomy in the text*

Factor or bias risk	1	0
Verification bias	n	y or ?
Disease progression bias	n	y or ?
Withdrawal bias	n	y or ?
Blinding biases	n	y or ?
Gold standard	100% pathology	< 100% pathology
Model	R only	L or R + L
Frequency	12 MHz	Not 12 MHz
Stenosis	n or ?	y
No. patients	≥ 100	< 100
Year of study	1990s	1980s

significance when testing multiple variables. Where n tests are performed, it is recommended[35] that the desired significance level is divided by n to arrive at the uncorrected probability that should be used to determine statistical significance, in this case a value of $p < 0.005$ would be required. This is sometimes known as the Bonferroni correction. No variables met this new stricter significance criterion. The 5% level of significance was used and the variables that individually met a 5% significance level were analysed. These results must be interpreted cautiously because the correction has not been applied; they are reported next for each of the six topic areas, three tumour staging and three lymph node staging, in the same order as in chapter 6.

7.1 Oesophageal tumour staging

Table 7.2 shows the results extracted from the 13 oesophageal studies. Only **patient number** had the 5% level of significance ($p = 0.033$). The studies were separated into two sets, one for the results with greater than 100 patients and one for those with less than 100 patients, and the SROC curve replotted based on the regression analysis. The gradient for each set is maintained as shown in

TABLE 7.2 *Factors and biases for oesophageal tumour staging*

Author and year of publication [reference]	Verification bias	Disease progression	Withdrawal bias	Blinding biases	Gold standard (pathology)	Model	Frequency	No.	Stenosis	Years
Binmoeller *et al.* 1995[3]	y	?	n	y	100%	R	7.5	< 100	y	1990s
Botet *et al.* 1991[4]	n	?	n	n	100%	R	7.5/12	< 100	y	1980s
Catalano *et al.* 1994[7]	y	n	n	n	100%	R	7.5/12	= 100	n	1990s
Dittler *et al.* 1993[9]	y	?	n	y	< 100%	R	7.5/12	> 100	y	1990s
Grimm *et al.* 1993[14]	y	?	n	y	< 100%	R	7.5/12	< 100	y	1990s
Heintz *et al.* 1991[17]	y	?	n	y	< 100%	R	7.5/12	< 100	n	1990s
Hünerbein *et al.* 1996[19]	y	?	n	y	100%	L	5/7.5	< 100	y	1990s
Manzoni 1993[20]	n	?	y	n	100%	L	7.5	< 100	n	1990s
Murata *et al.* 1988[22]	y	?	y	y	100%	R	7.5/10	> 100	n	1980s
Murata *et al.* 1993[23]	n	?	y	y	100%	R + L	7.5/12/20	> 100	n	1980s
Peters *et al.* 1994[26]	n	n	y	y	100%	R	7.5/12	< 100	n	1990s
Takemoto *et al.* 1986[29]	y	?	n	y	100%	R + L	7.5/12	< 100	y	1980s
Ziegler *et al.* 1991[35]	y	n	n	y	< 100%	L	7.5	< 100	y	1980s

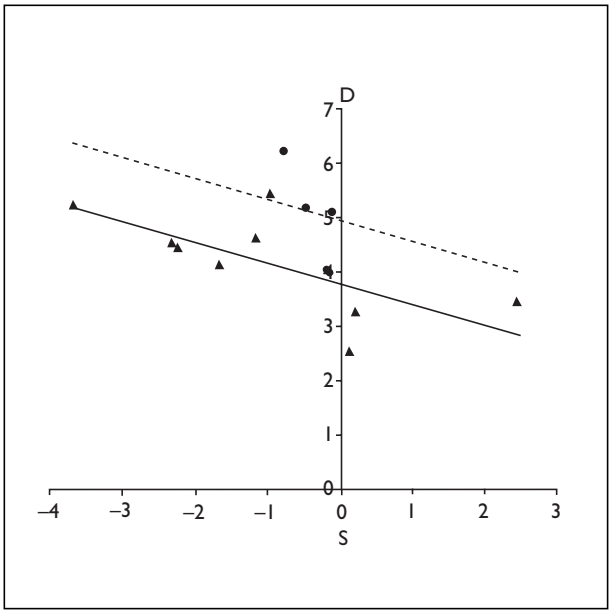


FIGURE 7.1 *SROC curves of patient number – oesophageal tumour staging (▲, D < 100; ●, D ≥ 100; ——, < 100; - - - -, ≥ 100)*

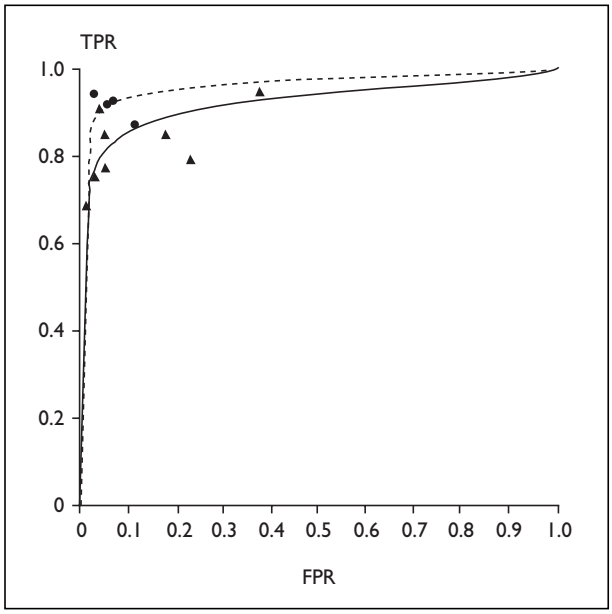


FIGURE 7.2 *ROC curves showing the influence of patient number – oesophageal tumour staging (▲, < 100; ●, ≥ 100; ——, < 100; - - - -, ≥ 100)*

Figure 7.1, but the intercept of each set is different *(Table 7.3),* which gives an indication of the log OR. *Figure 7.2* shows the transformation to ROC space and a clearer representation of the effect of patient number on the accuracy of staging oesophageal cancer can be seen from the curves.

If all the data from all the studies taken at face value were replotted on the SROC curve or the subsequent ROC curve, the line would lie between the two shown in *Figures 7.1* and *7.2*. This illustrates that one set (≥ 100 patients) represents an increase in the performance while the other represents a

TABLE 7.3 *Influence of patient number on oesophageal tumour staging*

Method	B	A	Se(A)	Q*
Total (n = 13)	–0.30	4.23	0.30	0.89
≥ 100 (n = 4)	–0.38	4.98	0.39	0.92
< 100 (n = 9)	–0.38	3.80	0.30	0.87
See section 3.1.5 for method				

decrease in the performance of EUS, with respect to the total data set.

7.2 Gastric tumour staging

The same factors and biases with the same dichotomy as oesophageal cancer studies were used in this analysis, as shown in *Table 7.1*. The results are shown for gastric cancer in *Table 7.4*.

When using the same significance level ($p < 0.05$), only the **type of probe used** was significant ($p = 0.029$). *Table 7.5* gives the results of the two fits for radial and linear probes. Results appear in *Figures 7.3* and *7.4*.

TABLE 7.5 *Influence of type of probe on gastric cancer staging*

Method	B	A	Se(A)	Q*
Total (n = 13)	0.80	5.26	0.29	0.93
Radial (n = 12)	0.57	5.35	0.24	0.94
Linear (n = 1)	0.57	3.21	0.84	0.83
See section 3.1.5 for method				

7.3 Gastro-oesophageal tumour staging

After combining all results of the three anatomical locations (oesophagus, stomach, and the cardia or gastro-oesophageal junction), an overall analysis was performed. The different biases and factors between the studies are shown in *Tables 7.2* and *7.4* for the oesophagus and stomach respectively. Only the additional factors and biases for the remaining studies addressing the accuracy of staging cardia or gastro-oesophageal junction cancer are supplied in *Table 7.6*. In this overall analysis the influence of the anatomical location on the staging accuracy was also evaluated.

TABLE 7.4 *Factors and biases for gastric tumour staging studies*

Author and year of publication [reference]	Verification bias	Disease progression	Withdrawal bias	Blinding biases	Gold standard (pathology)	Model	Frequency	No.	Stenosis	Years
Akahoshi *et al.* 1991[1]	n	?	n	y	100%	R	7.5/12	< 100	?	1980s
Botet *et al.* 1991[5]	n	?	n	n	100%	R	7.5/12	< 100	?	1980s
Caletti *et al.* 1993[6]	n	?	y	n	< 100%	R	7.5/12	< 100	?	1980s
Dittler *et al.* 1993[10]	y	?	n	y	< 100%	R	7.5/12	> 100	?	Both
Grimm *et al.* 1993[14]	y	?	y	y	< 100%	R	7.5/12	> 100	y	1990s
Hünerbein *et al.* 1996[19]	y	?	n	y	100%	L	5/7.5	< 100	y	1990s
Massari *et al.* 1996[21]	y	?	n	y	100%	R	7.5/12	< 100	?	1990s
Murata *et al.* 1988[22]	?	?	n	y	100%	R	7.5/10	> 100	n	1980s
Perng *et al.* 1996[25]	n	n	n	y	100%	R	7.5/12	< 100	?	1990s
Saito *et al.* 1991[27]	n	?	n	y	100%	R	7.5/12	> 100	?	1980s
Shimizu *et al.* 1994[28]	n	?	n	y	100%	R	7.5/12	> 100	?	1990s
Tio *et al.* 1989[30]	y	n	n	y	100%	R	7.5/10/12	< 100	?	1980s
Ziegler *et al.* 1993[36]	n	n	n	y	100%	R	7.5/12	>100	?	1980s
Both, 1980s and 1990s										

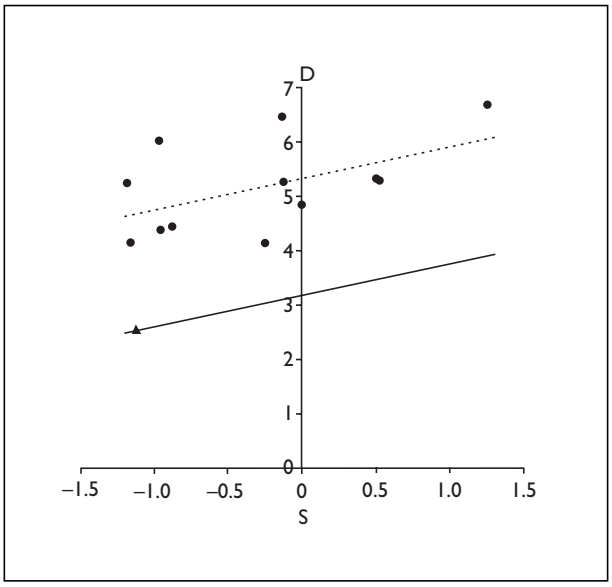


FIGURE 7.3 *SROC curves for type of probe – gastric tumour staging (▲, linear; ●, radial; ——, linear; - - - -, radial)*

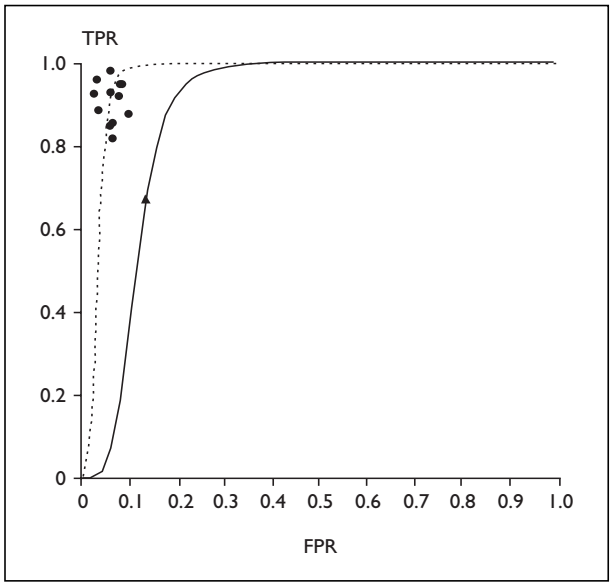


FIGURE 7.4 *ROC curves showing the influence for type of probe – gastric tumour staging (▲, linear; ●, radial; ——, linear; - - - -, radial)*

From the regression analysis, using the significance level of 5% ($p < 0.05$), three factors were significant *(Figures 7.5–7.10)*. These were the **year of the study** ($p = 0.011$), the **anatomical location** ($p = 0.016$), and the **presence of stenosis** ($p = 0.040$). *Figures 7.5, 7.7* and *7.9* show the SROC curves for these factors respectively; *Figures 7.6, 7.8* and *7.10* show the ROC curves for each set respectively. *Tables 7.7–7.9* show the results from the graphs.

As more than one factor was significant for gastro-oesophageal cancer tumour staging, the influence of putting a combination of factors into the regression model was assessed. The combination of year and anatomical location was significant ($p = 0.05$). It was hypothesised that the observed trend within the year of study could possibly be explained by the recent introduction of studies of the cardia. A further analysis was performed to investigate the influence of cardia studies on the relationship with the year of study. Studies investigating oesophageal or gastric cancer were combined together and compared with the cardia studies; the division into year of study was made on these new subsets. The results are shown in *Figures 7.11* and *7.12*, and *Table 7.10*. The difference between cardia 1990s and gastric and oesophageal 1980s studies was the most significant ($p = 0.0058$).

7.4 Lymph node staging of primary oesophageal tumours

For the analysis of staging accuracy of lymph nodes, the factor **size** was evaluated. For staging lymph nodes, many EUS image characteristics have been identified, which can be used to predict malignancy. These include echogenicity, homogeneity, border definition and shape. Although most studies

TABLE 7.6 *Factors and biases for cardia tumour staging studies*

Author and year of publication [reference]	Verification bias	Disease progression	Withdrawal bias	Blinding biases	Gold standard (pathology)	Model	Frequency	No.	Stenosis	Years
Altorki *et al.* 1996[2]	n	?	n	y	100%	R	7.5/12	< 100	y	1990s
François *et al.* 1996[11]	y	?	n	y	100%	R	7.5/12	< 100	y	1990s
Greenberg *et al.* 1994[13]	y	?	n	y	100%	R	7.5/12	< 100	y	1990s
Hordijk *et al.* 1993[18]	y	n	n	y	100%	R	7.5/12	< 100	y	1990s

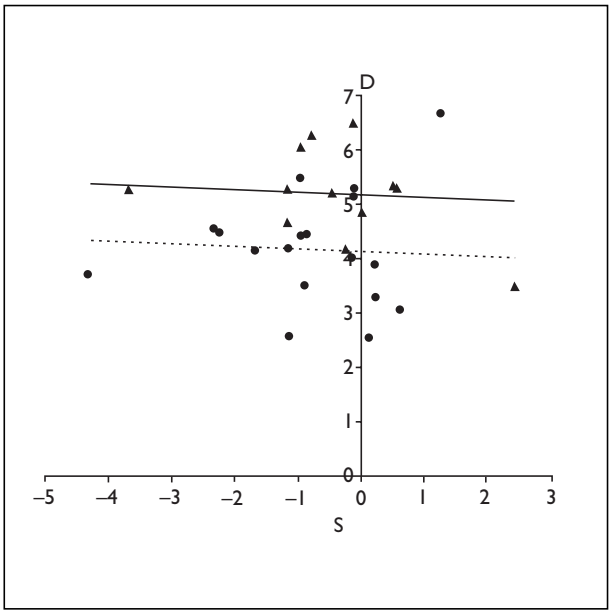


FIGURE 7.5 *SROC curves for year of study – gastro-oesophageal tumour staging (▲, 1980s; ●, 1990s; ——, 1980s; - - - -, 1990s)*

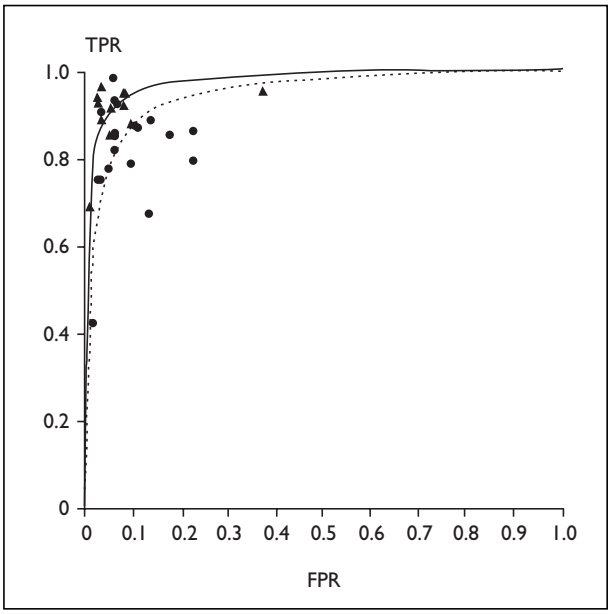


FIGURE 7.6 *ROC curves showing the influence of year of study – gastro-oesophageal tumour staging (▲, 1980s; ●, 1990s; ——, 1980s; - - - -, 1990s)*

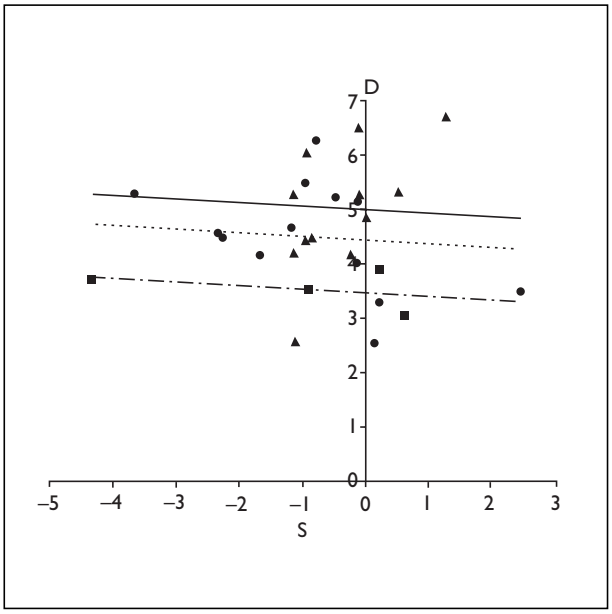


FIGURE 7.7 *SROC curves of anatomical location – gastro-oesophageal tumour staging (▲, stomach; ●, oesophagus; ■, cardia; ——, stomach; - - - -, oesophagus; – · – · –, cardia)*

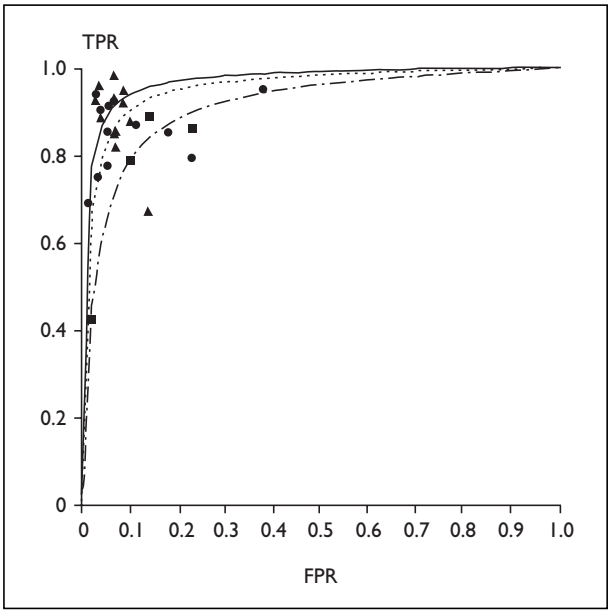


FIGURE 7.8 *ROC curves showing the influence of anatomical location – gastro-oesophageal tumour staging (▲, stomach; ●, oesophagus; ■, cardia; ——, stomach; - - - -, oesophagus; – · – · –, cardia)*

agreed on the use of these predictors, a small proportion reported that another predictor, the size of the lymph nodes, was also used as a diagnostic criterion. Hence, the additional influence of using size as a predictor of lymph node malignancy was assessed. The dichotomy of these factors and biases was the same as shown in *Table 7.1,* together with 'size not used = 1' and 'size used = 0'. In addition, stenosis was removed from the list shown in *Table 7.1* because this criterion was deemed to be less relevant in staging lymph nodes. *Table 7.11* shows the assigned values for each study.

From the regression analysis of the factors and biases, none was below the 5% significance level, although **blinding biases** had a significance level of $p = 0.054$. Further analysis was not performed.

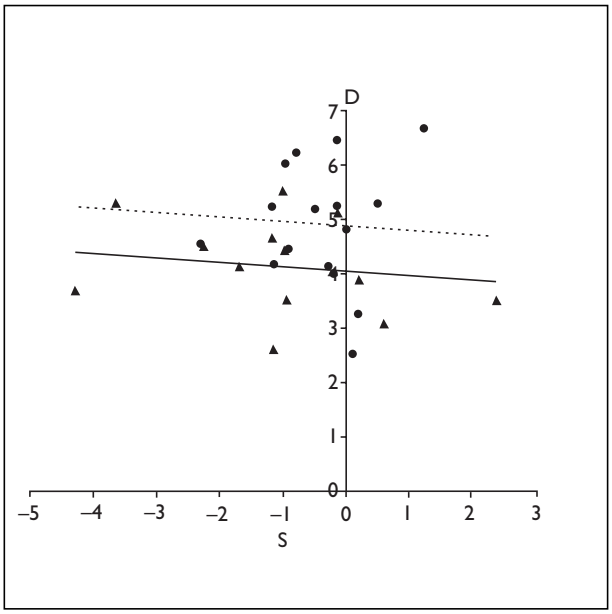


FIGURE 7.9 *SROC curves of stenosis – gastro-oesophageal tumour staging (▲, stenotic; ●, passable; ——, stenotic; - - - -, passable)*

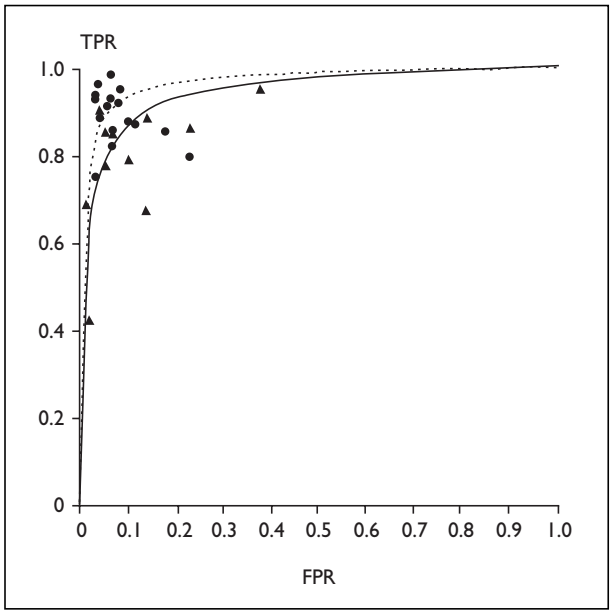


FIGURE 7.10 *ROC curves showing the influence of stenosis – gastro-oesophageal tumour staging (▲, stenotic; ●, passable; ——, stenotic; - - - -, passable)*

TABLE 7.7 *Influence of year of study on gastro-oesophageal tumour staging*

Method	B	A	Se(A)	Q*
Total (n = 30)	0.01	4.58	0.22	0.91
1990s (n = 18)	–0.05	4.13	0.32	0.89
1980s (n = 12)	–0.05	5.15	0.29	0.93
See section 3.1.5 for method				

TABLE 7.8 *Influence of anatomical location on gastro-oesophageal tumour staging*

Method	B	A	Se(A)	Q*
Total (n = 30)	0.01	4.58	0.22	0.91
Stomach (n = 13)	–0.07	4.96	0.29	0.92
Oesophagus (n = 13)	–0.07	4.42	0.31	0.90
Cardia (n = 4)	–0.07	3.45	0.53	0.85
See section 3.1.5 for method				

TABLE 7.9 *Influence of stenosis on gastro-oesophageal tumour staging*

Method	B	A	Se(A)	Q*
Total (n = 30)	0.01	4.58	0.22	0.91
Passable (n = 17)	–0.08	4.88	0.25	0.92
Stenotic (n = 13)	–0.08	4.03	0.33	0.88
See section 3.1.5 for method				

7.5 Lymph node staging of primary gastric tumours

The factors and biases analysed are shown in *Table 7.12*. None of these variables was significant at the 5% significance level, and therefore no further analysis was performed.

7.6 Lymph node staging of primary gastro-oesophageal tumours

For the differences between the studies in terms of factors and biases, *Tables 7.11* and *7.12* show the results for the oesophagus and the stomach respectively; *Table 7.13* shows the cardia results.

For this analysis, the influence of excluding outlying studies was also assessed.

First, for the total data set, only blinding bias was significant ($p = 0.019$). *Figure 7.13* shows the SROC curve divided by this bias; *Table 7.14* shows the results of the fits and Q*; and *Figure 7.14* shows the transformed ROC curves and illustrates the difference between studies with and without risk of blinding biases.

Secondly, for the data set with outlying studies excluded, only blinding bias was significant, but at a lower level ($p = 0.047$). *Figure 7.15* shows the SROC curve divided by this bias; *Table 7.15* shows

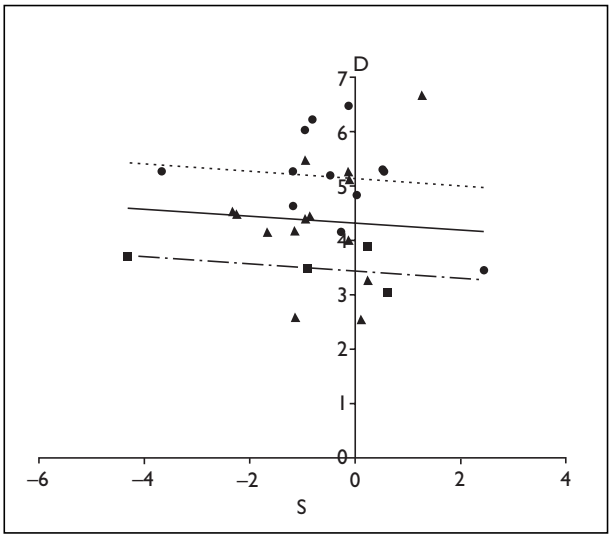


FIGURE 7.11 *SROC curves of year and anatomical location – gastro-oesophageal tumour staging (▲, 1990s: stomach and oesophagus; ●, 1980s: stomach and oesophagus; ■, 1990s: cardia; ——, 1990s: stomach and oesophagus; - - - -, 1980s: stomach and oesophagus; – · – · –, 1990s: cardia)*

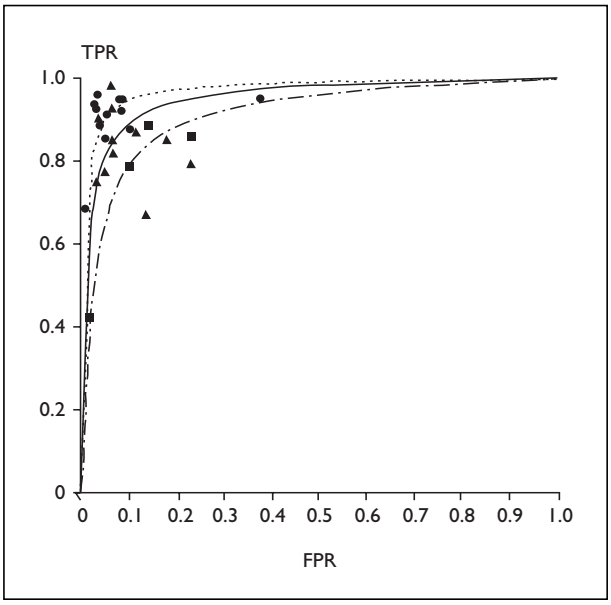


FIGURE 7.12 *ROC curves showing the influence of year and anatomical location – gastro-oesophageal tumour staging (▲, 1990s: stomach and oesophagus; ●, 1980s: stomach and oesophagus; ■, 1990s: cardia; ——, 1990s: stomach and oesophagus; - - - -, 1980s: stomach and oesophagus; – · – · –, 1990s: cardia)*

TABLE 7.10 *Influence of year and anatomical location on gastro-oesophageal tumour staging*

Method	B	A	Se(A)	Q*
Total (n = 30)	0.01	4.58	0.22	0.91
Stomach and oesophagus 1990s (n = 14)	–0.06	4.31	0.28	0.90
Stomach and oesophagus 1980s (n = 12)	–0.06	5.14	0.28	0.93
Cardia 1990s (n = 4)	–0.06	3.45	0.50	0.85
See section 3.1.5 for method				

the results of the fits and Q*; and *Figure 7.16* shows the transformed ROC curves and illustrates the difference between studies with and without risk of blinding biases.

Conclusion

To conclude, *Table 7.16* summarises the Q* values for all the subcategories. These values are discussed in chapter 8.

TABLE 7.11 *Factors and biases for lymph node staging studies associated with oesophageal tumours*

Author and year of publication [reference]	Verification bias	Disease progression	Withdrawal bias	Blinding biases	Gold standard (pathology)	Model	Frequency	No.	Size	Years
Binmoeller *et al.* 1995[3]	y	?	n	y	100%	R	7.5	< 100	y	1990s
Botet *et al.* 1991[4]	n	?	n	n	100%	R	7.5/12	< 100	y	1980s
Catalano *et al.* 1994[7]	y	n	n	n	100%	R	7.5/12	= 100	n	1990s
Dittler *et al.* 1993[9]	y	?	n	y	< 100%	R	7.5/12	> 100	y	Both
Grimm *et al.* 1993[14]	y	?	n	y	< 100%	R	7.5/12	< 100	y	Both
Heintz *et al.* 1991[17]	y	?	n	y	< 100%	R	7.5/12	< 100	y	Both
Hünerbein *et al.* 1996[19]	y	?	n	y	100%	L	5/7.5	< 100	y	1990s
Peters *et al.* 1994[26]	n	n	y	y	100%	R	7.5/12	< 100	n	1990s
Ziegler *et al.* 1991[35]	y	n	n	y	< 100%	L	7.5	< 100	y	1980s
Both, 1980s and 1990s										

TABLE 7.12 *Factors and biases for lymph node staging studies associated with gastric tumours*

Author and year of publication [reference]	Verification bias	Disease progression	Withdrawal bias	Blinding biases	Gold standard (pathology)	Model	Frequency	No.	Size	Years
Botet *et al.* 1991[5]	n	?	n	n	100%	R	7.5/12	< 100	y	1980s
Dittler *et al.* 1993[10]	y	?	n	y	< 100%	R	7.5/12	> 100	y	Both
Grimm *et al.* 1993[14]	y	?	y	y	< 100%	R	7.5/12	> 100	y	Both
Hünerbein *et al.* 1996[19]	y	?	n	y	100%	L	5/7.5	< 100	y	1990s
Massari *et al.* 1996[21]	y	?	n	y	100%	R	7.5/12	< 100	y	1990s
Perng *et al.* 1996[25]	n	n	n	y	100%	R	7.5/12	< 100	n	Both
Tio *et al.* 1989[30]	y	n	n	y	100%	R	7.5/10/12	< 100	y	1980s
Ziegler *et al.* 1993[36]	y	n	n	y	100%	R	7.5/12	> 100	y	Both
Both, 1980s and 1990s										

TABLE 7.13 *Factors and biases of lymph node staging of primary tumours at the cardia*

Author and year of publication [reference]	Verification bias	Disease progression	Withdrawal bias	Blinding biases	Gold standard (pathology)	Model	Frequency	No.	Size	Years
Altorki *et al.* 1996[2]	n	?	n	y	100%	R	7.5/12	< 100	y	Both
François *et al.* 1996[11]	y	?	n	y	100%	R	7.5/12	< 100	y	1990s
Greenberg *et al.* 1994[13]	y	?	n	y	100%	R	7.5/12	< 100	y	1990s
Both, 1980s and 1990s										

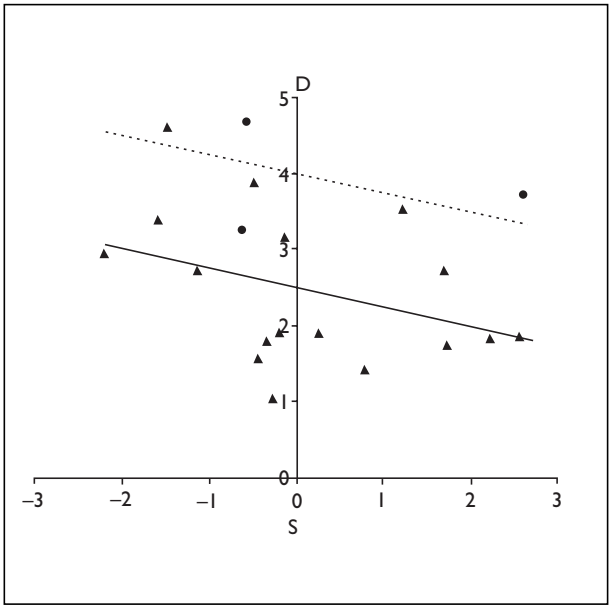


FIGURE 7.13 *SROC curves of blinding biases – lymph node staging of primary gastro-oesophageal tumours: total data set (▲, bias; ●, no bias; ——, bias; - - - -, no bias)*

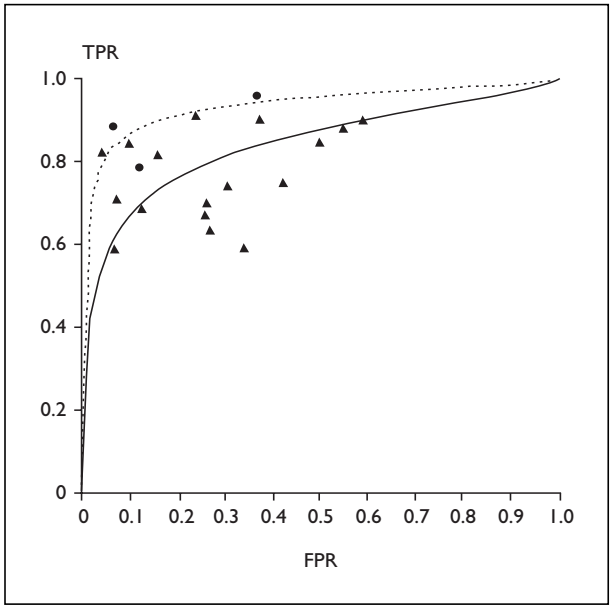


FIGURE 7.14 *ROC curves showing the influence of blinding biases – lymph node staging of primary gastro-oesophageal tumours: total data set (▲, bias; ●, no bias; ——, bias; - - - -, no bias)*

TABLE 7.14 *Influence of blinding biases on lymph node staging of primary gastro-oesophageal tumours: total data set*

Method	B	A	Se(A)	Q*
Total (n = 20)	–0.22	2.71	0.24	0.79
No bias (n = 3)	–0.26	4.00	0.54	0.88
Biased (n = 17)	–0.26	2.49	0.22	0.78
See section 3.1.5 for method				

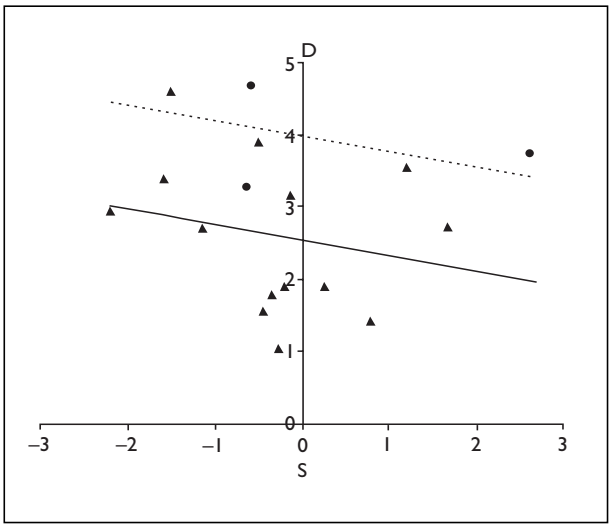


FIGURE 7.15 *SROC curves of blinding biases – lymph node staging of primary gastro-oesophageal tumours: data set with outlying results excluded (▲, bias; ●, no bias; ——, bias; - - - -, no bias)*

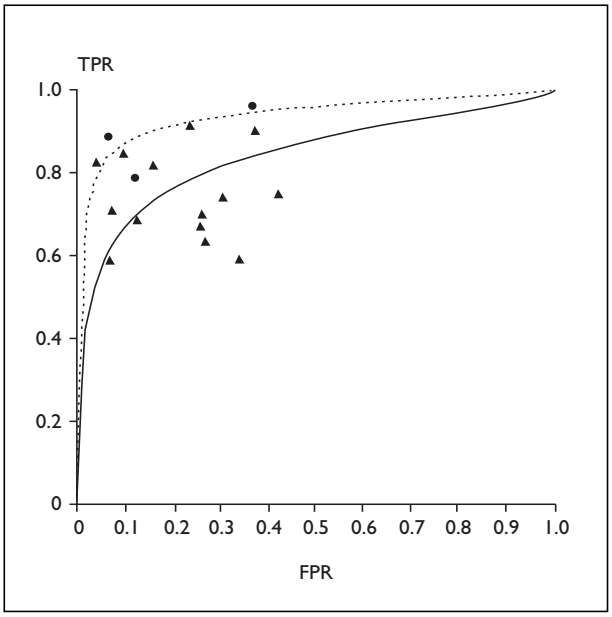


FIGURE 7.16 *ROC curves showing the influence of blinding biases – lymph node staging of primary gastro-oesophageal tumours: data set with outlying results excluded (▲, bias; ●, no bias; ——, bias; - - - -, no bias)*

TABLE 7.15 *Influence of blinding biases on lymph node staging of primary gastro-oesophageal tumours: data set with outlying results excluded*

Method	B	A	Se(A)	Q*
Total (n = 17)	–0.10	2.81	0.28	0.80
No bias (n = 3)	–0.22	3.93	0.59	0.88
Biased (n = 14)	–0.22	2.53	0.28	0.78
See section 3.1.5 for method				

TABLE 7.16 *Summary of Q* values for each topic*

Topic	Q*
Oesophageal tumour staging	
Full data set (n = 13)	0.89
≥ 100 patients (n = 4)	0.92
< 100 patients (n = 9)	0.87
Gastric tumour staging	
Full data set (n = 13)	0.93
Radial probes (n = 12)	0.94
Linear probes (n = 1)	0.83
Gastro-oesophageal tumour staging	
Full data set (n = 30)	0.91
1990s (n = 18)	0.89
1980s (n = 12)	0.93
Stomach (n = 13)	0.92
Oesophagus (n = 13)	0.90
Cardia (n = 4)	0.85
Passable (n = 17)	0.92
Stenotic (n = 13)	0.88
Stomach or oesophagus and 1990s (n = 14)	0.90
Stomach or oesophagus and 1980s (n = 12)	0.93
Cardia 1990s (n = 4)	0.85
Lymph node staging of primary oesophageal tumours	
Full data set (n = 9)	0.82
No significant variables	–
Lymph node staging of primary gastric tumours	
Full data set (n = 8)	0.76
No significant variables	–
Lymph node staging of primary gastro-oesophageal tumours	
Full data set (n = 20)	0.79
No bias (n = 3)	0.88
Blinding biases (n = 17)	0.78
Data set with outlying results excluded (n = 17)	0.80
No bias (n = 3)	0.88
Blinding biases (n = 14)	0.78

Chapter 8
Discussion

In this chapter the review methodology is discussed, with particular reference to potential biases in the approach. The results of the review reported in chapters 4 and 6, analysed for robustness in chapter 7, are discussed with reference to the questions posed in chapter 2. This chapter concludes with an overview of the development of the literature in this field.

8.0 Review methodology

8.0.1 Search strategy

The search strategy was designed to have a high recall and therefore low precision. No limitations on publication type, such as the search for RCTs,[12] were applied owing to the dearth of data available. As the MeSH term 'endosonography' came into use only after 1993 and no such classification is available for BIDS, alternative keyword searching was necessary to identify EUS studies. Because of the diversity of EUS technology and the range of terminology that is used medically, no attempt was made to search electronically the specific topic of staging gastro-oesophageal cancer. Instead, the abstracts of those studies retrieved from the keyword search for EUS were systematically read and classified. This system is labour intensive and requires the full publication to be assessed if no abstract is provided. However, the system used had the following advantages.

- It does not rely on the accuracy of MeSH terms.
- It is still efficient if no subject classifications are available.
- It allows the use of as many individually designed classifications as required.
- Once the process is complete, if a database of classifications is maintained, a record of both included and excluded data, together with reasons, can easily be accessed and searched.
- Because no data are discarded, extension of the coverage of the review will not require any further searches to be performed.

When searching for evidence on the higher levels of the hierarchical framework, the same process of using suitable keywords was applied. Although the searches were more general than at the staging performance level because no specific area of clinical application was considered, more MeSH terms, such as 'outcome and process assessment (health care)', were used to increase the precision of the search for studies at the required levels of the Fineberg framework.

In contrast to the findings of other workers, the bulk of the studies identified were in MEDLINE. For example, Dickersin *et al.*[36] reported the sensitivity of MEDLINE to be as low as 51% for RCTs in ophthalmology. Of the 27 EUS staging studies included in this review, MEDLINE identified 24 (89%). In addition, in our review of the health economics of EUS, MEDLINE supplied all the information. On reanalysing the occurrence of the keywords and MeSH terms used in the search, it was found that the MeSH term 'economics' was sufficient to retrieve the two studies we identified as the best available evidence. Indeed, this finding was repeated in another review.[17]

Of the three staging studies not retrieved from MEDLINE, one was found by searching the cited reference lists of retrieved articles, one was retrieved from Inside Information Plus (British Library),[14] and one from FirstSearch (OCLC).[15] The additional impact of these studies on the results was not assessed, so the potential benefit of searching these resources cannot be quantitatively reported. However, findings using the checklist and the proposed scoring showed that the quality of these three studies was comparable with the best of the studies identified by MEDLINE.

Regarding publication bias, our search for any information on unpublished and grey literature was unsuccessful. There are three possible explanations for this.

- The poor response rate to requests for information may have occurred owing to authors' unwillingness to have negative or equivocal results publicised.
- Our directed mailings were not reaching the correct targets.
- The information retrieved was in fact representative – publication bias is not a problem in this subject area.

8.0.2 Inclusion criteria

The quality of a review depends explicitly on the quality of the primary studies. The inclusion criteria

must be based on the level of evidence available. For example, setting an inclusion criterion of an RCT design for this review would not have been profitable. The type of study available for this review led to inclusion criteria that were primarily not concerned with the study design. The criteria chosen were those sufficient to allow adequate comparison of studies in terms of the presentation of results as well as the field of the study. For the studies evaluating the higher levels of the Fineberg framework, the criteria were less stringent owing to the lack of data. The criteria were set to include studies that supplied any information, irrespective of the study design or quality.

The decision to exclude non-English language studies was probably the largest potentially biasing influence in the design of our review. This decision was made for pragmatic reasons. The initial intention was not to exclude studies on the basis of language, but the searches returned a larger number of non-English language articles and a wider range of languages than had been envisaged. It was decided to restrict the review to English language publications rather than translate a possibly unrepresentative subset of papers, as it would not have been possible to translate all the non-English language papers within the time and budget available, even if technical translators for the full range of languages had been located.

There is a possibility that our inclusion criteria could have led to the exclusion of studies with negative or equivocal results. If it is hypothesised that information is more likely to be omitted from studies with such findings, then our requirement for the publication of numerical results suitable for inclusion in a 2×2 contingency table could lead to the preferential inclusion of studies with positive findings. This selectivity would have an effect similar to that of publication bias.

8.0.3 Assessment of relevance and validity of primary studies

In many of the results discussed in chapter 7 it was stated that there was no statistically significant relation between the presence of bias and the study result. This finding is not as conclusive as one might hope owing to missing data in the regression analysis. In common with other authors we chose to describe a risk of bias as present if the information required to determine its presence was not presented in the paper. Omissions in the descriptions of the primary studies were so widespread that it was not possible to contact the authors to confirm their exact methodology. There are two possible outcomes.

- A proportion of studies where there was no risk of bias have been classified as having the fault, and this will have prevented any relationship being significant at the specified level.
- The bias risk was present in almost all the studies, meaning that there were insufficient studies without the bias risk for discrimination in the analysis.

The most prevalent bias risks apparent in the primary studies were blinding biases and verification bias. The risk of verification bias was not found to be significantly related to the study results in any of the analyses. Only for lymph node staging was the blinding of results found to influence the study results; this is discussed further in section 8.1.

It could be argued that it is not surprising that different studies achieved different results if they used, for example, different equipment or investigated completely different sets of patients (these are examples of the variables we have called factors). We chose to include all studies in the quantitative analysis, in spite of knowing that their results would possibly be different because of the wide range of studies described in the literature and the likelihood of being unable to perform any synthesis of results without the regression approach. This had the additional advantage of giving quantitative evidence of any differences.

8.0.4 Data extraction

Our review methodology was designed to minimise bias by using a multidisciplinary panel from more than one centre. After an initial phase to check inter-reviewer reproducibility, the data extraction was performed primarily by a single reviewer who consulted other panel members if in doubt. It is possible that this strategy may have introduced a degree of bias to our review process, but it is not believed to be large because, in the event, the more difficult and potentially subjective decisions (such as presence of bias risk) were not used as inclusion criteria.

8.0.5 Quantitative data synthesis

The methodology used to synthesise the data from the studies into an SROC curve was proposed by Moses *et al.*[25] and developed by Irwig *et al.*[26] Still in its developmental stages, meta-analysis applied to diagnostic tests requires practical implementation to refine its use. Two aspects of the methodology were investigated in this review:

- the exclusion of points on an ROC scatterplot lying outside a range of clinical utility defined as sensitivity and specificity both greater than 50%[25]

- the fitting of a line to the points of the SROC curve, using either the EWLS method or the RR method.

For the first aspect, a complete analysis was obtained for the data set on the staging of gastro-oesophageal lymph nodes. Three studies were outside the defined range. For the full data set the resulting performance statistic was Q* = 0.79, compared with that of the data set with outlying results excluded, for which Q* = 0.80. Further work is needed in this area before conclusions may be drawn about exclusion. Blinding was found to be significant for both EWLS and RR methods but with differing *p*-values. On comparing potentially unbiased and biased studies for each method, no significant difference in Q* was found between the full data set and that with outlying results excluded.

For the second methodological aspect, both EWLS and RR methods were used to plot lines to the SROC curve and results of gradients and intercepts were compared. A clear difference between the methodologies was observed for all topics, but with the greatest difference occurring for the data set concerning the staging of oesophageal lymph nodes.

For lymph node staging of primary gastro-oesophageal tumours, the EWLS and the RR methods were used to plot the ROC curve, both for the total data set and the data set with outlying results excluded. Although the relative influence of the two methods on the Q* value was small, the shape of the ROC curve deteriorated slightly for the RR plot, with the curve forming an 'S' shape. This 'S' shape became more evident on the data set with outlying results excluded *(Figure 6.21)*. Also, unlike the RR plot, the EWLS method allowed the simple calculation of the Se for the fit of the line. For these reasons, especially if the number of included studies in the analysis is small, the EWLS plot was preferred and therefore incorporated into the regression analysis.

The choice of using Q* to summarise the ROC curve was considered relevant to this topic owing to the desired effect of balancing under-staging and over-staging, as argued in section 3.1.5. For a different clinical application of EUS, or for other tests, an alternative summary statistic may be a better indicator of performance. Such an estimate could be the TPR value read off the ROC curve from the mean FPR, which is perhaps a more generalisable indication of test performance. As the variation in threshold associated with EUS is implicit, although Q* is the optimal balance between sensitivity and specificity, not every operator will be able to attain this, and no explicit recommendation can be given on how to reach this value.

8.1 EUS in gastro-oesophageal cancer – staging performance

In order to address the broad question of whether EUS staging should be recommended for gastro-oesophageal tumours, a number of very specific questions were posed about the staging performance of EUS, looking individually at the T, N and M classifications and at the three anatomical sites: oesophagus, stomach and cardia. As discussed in section 3.1.5, we chose Q*, the value on the SROC curve where sensitivity equals specificity, as our summary statistic. The 'positive' classification is of the lower stage (section 3.1.4), so low specificity is associated with under-staging of tumours and low sensitivity with over-staging. As neither over-staging nor under-staging is desirable, the choice of a statistic that does not emphasise one at the expense of the other is optimal.

8.1.1 T staging

The staging performance of EUS for distinguishing Stages T1 and T2 from Stages T3 and T4 for **tumours of the oesophagus** is Q* = 0.89 (n = 13; *Figure 6.3)*. The most prevalent biases apparent in the primary studies were blinding biases and verification bias, but we did not find a significant ($p < 0.05$) relation between the presence of bias and the study results. The number of patients enrolled in the study had a significant effect, with those studies (n = 4) with ≥ 100 patients suggesting better staging performance (Q* = 0.92) than those with fewer (n = 9; Q* = 0.87; *Figure 7.2)*.

The staging performance of EUS for distinguishing stages T1 and T2 from Stages T3 and T4 for **tumours at the cardia** was not assessed numerically. Four primary studies were found, one addressing only tumours at the cardia, one including tumours at the cardia, and two including tumours of the gastro-oesophageal junction.

The staging performance of EUS for distinguishing Stages T1 and T2 from Stages T3 and T4 for **tumours of the stomach** is Q* = 0.93 (n = 13; *Figure 6.6),* which is the highest value found for any of the subcategories. Blinding biases were even more prevalent than for the studies of oesophageal tumours, and again we did not find a significant ($p < 0.05$) relation between the presence of the risk of bias and the study results. Although only one study was included that used a linear probe, the results from this study were significantly different ($p < 0.05$) from those for the rest of the

studies, which used radial probes. Better staging performance (n = 12; Q* = 0.94) was attained for radial probes than for the linear probe study (n = 1; Q* = 0.83; *Figure 7.4).* This is further discussed below with reference to the question about the use of dedicated or non-dedicated EUS equipment.

By combining the primary studies in the individual analyses above, the staging performance of EUS for distinguishing Stages T1 and T2 from Stages T3 and T4 for **tumours of the oesophagus, at the cardia and of the stomach** is Q* = 0.91 (n = 30; *Figure 6.9).* Again, although review and verification bias risks were noted for many of the included primary studies, we did not find a significant relation between the risk of the bias and the study results. The difference between the Q* values quoted for the three regions separately were shown to be statistically significant ($p < 0.05$), with EUS performing best in the stomach and least well at the cardia *(Figure 7.8).* Methodological differences between the primary studies also affected the summary results ($p < 0.05$). Some studies counted all those individuals who underwent the examination in the total number of patients, regardless of whether or not the tumour was traversable. Others defined as the total number in the study only those for whom the tumour was traversable. A result of Q* = 0.92 was calculated from traversable tumours only. A more realistic figure, if all the patients actually undergoing the procedure were included, was Q* = 0.88 *(Figure 7.10).* Failure to cross a stenosis is not a problem unique to the use of EUS in gastro-oesophageal applications (for example, it may arise in EUS of a large pancreatic tumour, or in the colon). The prevalence of the problem is of interest to those considering using the technique and it is discussed further below. The other statistically significant relations ($p < 0.05$) were all related to the date of the study. A higher Q* (0.93) was obtained for the 12 studies from the 1980s than for those of the 1990s (n = 18; Q* = 0.89; *Figure 7.6).* None of the earlier studies tackled tumours at the cardia, which are a more difficult staging challenge for both EUS, owing to the oblique angulation, and CT,[37] but even when the cardia studies are removed from the analysis there remains a significant difference between the two decades: Q* = 0.93 for the 1980s and 0.90 for the 1990s *(Figure 7.12).* Interpreting this finding is fraught with complicating issues, but other authors[38] have reported falls in the performance of imaging tests when they move from being applied to a well-defined research population by experts to more general application, which occurs with time.

8.1.2 N staging

The staging performance of EUS for distinguishing Stage N0 from Stage N1 for **lymph nodes associated with primary oesophageal tumours** is Q* = 0.82 (n = 9; *Figure 6.12).* We did not find a statistically significant ($p < 0.05$) relation between the presence of bias and the study results.

The staging performance of EUS for distinguishing Stage N0 from Stages N1 and above for **lymph nodes associated with primary tumours** at the cardia was not assessed numerically because only three primary studies were found.

The staging performance of EUS for distinguish-ing Stage N0 from Stages N1 and above for **lymph nodes associated with primary gastric tumours** is Q* = 0.76 (n = 8; *Figure 6.15).* This was the lowest value found for any of the subcategories and suggests that EUS is at its least effective for lymph node staging. We did not find a statistically significant ($p < 0.05$) relation between the risk of bias and the study results. For comparison, and it must be emphasised that these figures do not result from a systematic review of the literature, alternative methods do not perform any better. For example, for spiral CT the sensitivity for distinguishing Stage N0 from N1 was found to be 24%, with specificity of 100%.[39]

By combining the primary studies in the individual analyses above, the staging performance of EUS for distinguishing Stage N0 from Stages N1 and above for **lymph nodes associated with primary tumours of the oesophagus, at the cardia and of the stomach** is Q* = 0.79 (n = 20; *Figure 6.18).* This category was the only one where the risk of a bias in the primary study, in this case blinding biases, had a statistically significant ($p < 0.05$) effect on the Q* result. The finding has implications for the assessment of study validity, as one might argue that for these studies there is a quality indicator and that studies exhibiting risk of blinding biases should be excluded. It would be expected that the presence of blinding biases would lead observers to perform better than if they had not had access to the additional information. Our analysis *(Figure 7.14)* resulted in Q* = 0.88 for three studies where the risk of bias had been avoided, and Q* = 0.78 (n = 17) where the bias risk was present. In other words, there was better performance when the risk of bias was **not** present. We conclude that the results must be treated cautiously and that other confounding factors may be involved, although they were not found to be statistically significant. A possible reason is the widespread lack of reporting in the primary studies of information needed to assess study quality.

8.1.3 M staging

The staging of metastases (M staging) is a particularly complex part of the TNM classification. The

definition is slightly different for the oesophagus and stomach (chapter 1). The five primary studies are summarised in *Table 4.9*. There are too few studies in the separate anatomical areas to combine their numerical results. Although this lack of evidence would suggest a need for further research, this is not recommended. The limited depth of penetration of EUS makes it unsuitable as an imaging modality for the exclusion of distant metastases because the whole of the liver, the lungs and more distant lymph node groups cannot be visualised.

8.1.4 Grouped TNM staging

The grouped TNM staging system was described in *Figure 3.2b*. It is a less sensitive scheme because it combines the T, N and M stages into a smaller number of groupings; it is not widely used in the UK. We have included it for completeness and because some literature exists on the topic. Only five primary studies were found. The information in these articles was insufficient to allow assessment of the performance against a reference standard for the full classification scheme. Theoretically, the study that needs to be performed is one where the decisions based on the full TNM classification are compared with those based on the grouped stage. However, this is not one of our recommendations because one needs to have all the separate TNM stage classifications to define the grouped stage, so one may as well use the full classification scheme.

8.1.5 Unsuitability of patients due to non-traversability

This problem was touched on briefly above, where we noted that some primary studies included in their calculations only those patients for whom the procedure was fully completed, thus masking the well-known problem of a failed study due to non-traversable stenosis. In such studies the performance will appear better than in those where all patient results are included in the analysis. Eleven of 27 studies that met our inclusion criteria for staging performance stated the proportion of impassable stenoses for 13 patient groups. These proportions are illustrated in *Figure 8.1*. (Only those where no attempt at dilatation was made are included.) There is considerable variation in the proportion and it is not possible therefore to state a figure for the expected number of non-traversable tumours. We recommend that this is an area that should be subject to continued systematic review, with the aim being to determine more precisely the proportion of failed studies due to stenosis. A separate but related question was not addressed in this review, that of whether the performance of dilatation prior to evaluation by EUS should be recommended.

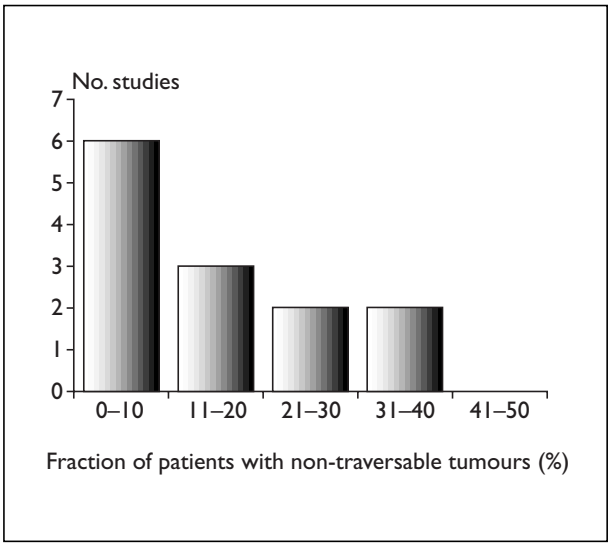


FIGURE 8.1 *Percentage of patients (on whom EUS was attempted) with a non-traversable tumour; includes studies of oesophageal, gastric and cardia tumour staging*

8.1.6 EUS miniprobes

The prime reason for using a miniprobe is for its ability to make subclassifications of T1 tumours as mucosal or submucosal. They are not intended to replace conventional probes as they cannot be used for conventional T, N or M staging owing to their poor depth of penetration. Three studies were found, two investigating oesophageal tumours, one gastric *(Table 4.6)*. The number of T1 tumours was small for each study and, although their findings appear to show better performance in the oesophagus than the stomach, the evidence is not strong. A role for the miniprobe is in determining the depth of tumour penetration at high spatial resolution prior to endoscopic mucosal resection, but it should not be used in isolation as it would not be possible to identify distant spread using the miniprobe alone. A further role for the miniprobe might be to overcome the non-traversability problem. In practice this is not a sensible role because its poor depth of penetration limits both N and M staging performance.

At this time there is no firm evidence about the value of miniprobes and we recommend further research in this area.

Conclusion

Overall then, EUS has been shown to be an accurate technique for staging gastro-oesophageal tumours, performing best for gastric tumours. Its performance for lymph node staging is less good and this must be judged in relation to

alternative modalities. There is insufficient evidence to draw conclusions about performance at the cardia, for the staging of metastases or of miniprobes.

8.1.7 Reproducibility and learning curve

The study by Catalano *et al.*[46] was less realistic in design than that of Burtin *et al.*[45] because of the reliance on video examination, but this did separate the operating and assessment functions of the examination. It was found[45] that intraobserver agreement was better where both assessments were made from a standardised recording, as opposed to one recording and one active examination. Thus the studies would be expected to give better reproducibility than is achievable in practice, without strict guidelines on a standardised protocol. It is not possible to combine the results of the studies numerically, but they had some conclusions in common. Both showed generally good agreement for T staging and T2 lesions showed the worst agreement, which corresponds to lower staging accuracy for this type of tumour. For N staging, Catalano *et al.*[46] found good agreement, Burtin *et al.*[45] found only satisfactory agreement for paraoeophageal nodes and intra-abdominal nodes; agreement was less good for paratracheal nodes. They point out that there is also a site dependence for both prognostic significance and alternative staging possibilities. The comparisons in Catalano *et al.*'s[46] study also have relevance to the learning curve question. These authors found a significant difference in intra- and interobserver reproducibility between experienced (> 50 examinations) and inexperienced (< 20 examinations) observers. The actual number of procedures performed by the experienced group is not known. Accuracy rates for the experienced group were 84–90% for T stage and 80–88% for N stage; for the inexperienced observers the corresponding figures were 54–66% (T) and 56–64% (N). Fockens *et al.*[47] addressed the question in terms of the staging accuracy of a single gastroenterologist and found a similar level of improved accuracy in the second period, from 58% to 83% for T staging. Both studies suggested that inexperienced ultrasonographers may make errors in balloon overinflation, which results in overstaging of T1 lesions. Fockens *et al.*[47] suggest that acceptable accuracy rates are achievable only after 100 examinations, although this is a somewhat arbitrary choice.

The information available regarding reproducibility and the learning curve is sparse. The studies recommend extended training and further studies of reproducibility as the technology is improved.

8.2 EUS in gastro-oesophageal cancer – staging impact

To address the question of whether there is any evidence that EUS has an impact on methods used for staging, four specific questions were posed.

8.2.1 CT/MRI/positron emission tomography compared with EUS

To be sure that our comparison of CT/MRI/positron emission tomography and EUS results was valid we sought studies that performed both imaging tests on the same set of patients and compared the results with the same gold standard. Only comparisons with incremental CT were found. This approach has the merit of ensuring comparability of the results but it can lead to biased findings for the following reasons. Many such comparative studies are performed by exponents of one of the included technologies. They bring to the study their preconceptions and, as experts in their favoured technology, may not perform the comparators with equal skill. In the eight studies comparing EUS with incremental CT for staging, the CT protocols were less well described than those for EUS and may not have been the best achievable. It is significant that only one team discussed the complementary roles of the two modalities, increasing the concern that the authors aimed to prove superiority of their favoured modality. The quality of the evidence is poor and we do not believe it appropriate to draw any conclusions from the primary studies, other than to draw on their shortcomings for our recommendations.

8.2.2 EUS miniprobes versus conventional EUS

Only one study in this category was found, but this is not surprising as the roles of the two types of probe are quite distinct, as outlined in section 8.1. An investigation of the complementary role of EUS miniprobes and conventional probes may be more suitable; however, the one study failed to evaluate this application.

8.2.3 EUS-guided FNA in gastro-oesophageal cancer

As outlined in chapter 1, this is an attractive application of EUS, allowing the fine-needle biopsy of lymph nodes to help in staging. However, validation of biopsy techniques is difficult, especially where sample sizes are small, as described in chapter 1. Seven primary studies were found but it is not possible to draw conclusions from the individually rather varied methodologies and mixtures of tumour types.

8.2.4 Dedicated EUS systems

The newcomer to the literature would not in general be able to determine whether the probe in use was one that could be connected to an existing ultrasound machine, or was part of a dedicated EUS system. A general rule is that radial probes are associated with dedicated systems although linear probes are not. We found no primary studies performing a direct comparison of the two types of probe on the same group of patients. There is some indirect evidence, which is discussed in section 8.1 above, in relation to the staging performance of EUS for distinguishing Stages T1 and T2 from Stages T3 and T4 for tumours in the stomach. The finding of significantly better performance from radial probes (illustrated in *Figure 7.4)* was based on results that included only one study using a linear probe, and so should be further investigated. The level of evidence is currently insufficient for any recommendations on purchasing to be made.

8.3 EUS in any clinical application – therapeutic impact

Four studies were found that addressed the question of the therapeutic impact of EUS; they are described in section 4.3. Only one study of oesophageal cancer was found that met the inclusion criteria. Three others were included relating to mixed anatomical areas. The main weakness of these studies was test review bias. There were no randomised comparisons of EUS therapeutic impact that empirically tested therapeutic decisions with and without EUS. Most designs used sequential testing of the same patient group, with questionnaires being completed by physicians or ultrasonographers after each test result. Therapeutic decisions were ultimately made with the results of all tests known. Three of the studies claimed positive therapeutic impact for EUS. It is interesting that the only study using independent test review showed no therapeutic impact, but this study suffered from other biases, particularly in patient selection (Fok *et al.*[37]).

The study designs and clinical applications were varied, so firm conclusions may not be drawn. The general impression was that EUS could have an important effect on therapeutic decisions regarding surgical intervention, by either preventing unnecessary surgery in advanced disease or indicating the possibility of more aggressive treatment by better staging of less advanced disease.

However, the study designs used were generally weak, particularly with regard to referral bias and test review bias. Identification of therapeutic impact is an important stage in determining outcomes and cost-effectiveness. Because of the lack of controlled, comparative studies in this area it is not possible to judge the independent influence of EUS on therapeutic decisions.

8.4 EUS in any clinical application – patient outcome

There is a definite lack of evidence in this area. Only two studies were found that addressed the question of the effect of EUS on patient outcome in a systematic way. They are described in section 4.4. The studies were not designed specifically to assess outcomes and did not approach the question in the manner suggested in chapter 1. The outcome measures used were basic clinical endpoints such as survival. In each study the outcome data were incomplete with follow-up on only a subsample of patients. Two other studies were found, which included useful data on survival after surgery by disease stage but, as there was no comparative element, the impact of EUS as opposed to other techniques could not be judged. The data may however be useful in future model-based studies of outcome and cost-effectiveness.

8.5 EUS in any clinical application – health economics

From the 32 English language articles located in the review, only two merited further consideration and only one of these was a study designed specifically to address economic issues (Allgayer *et al.*[43]). This study also addressed the clinical question of interest, namely the staging of gastrointestinal tumours, comparing the cost-effectiveness of EUS and CT scanning. Unfortunately, the study was weak in all the four key areas identified in section 3.5.3. The study design was retrospective rather than prospective. The data collection was restricted to costs of the diagnostic investigations and the effectiveness indicator was the intermediate clinical endpoint, 'correct diagnosis'. Neither the actual costs of the investigations nor their sensitivity were directly observed in the study patients. Costs were taken from procedure reimbursement rates in Germany and sensitivities were derived from the literature. No attempt was made to use data on patient outcomes in the effectiveness measurement. The study might still have been useful if the analysis had been correctly conducted and presented. However, there is confusion over the use of cost-effectiveness ratios. EUS is in fact the dominant

strategy, being both less costly and more effective in diagnosis. The authors did not appreciate that in such circumstances the presentation of a cost-effectiveness ratio is neither possible nor necessary. The second study (Prat *et al.*[44]) addressed the choice of diagnostic investigation for suspected bile duct lithiasis. The study design was prospective but not randomised. Patients were given both EUS and ERC by independent investigators, followed by endoscopic sphincterotomy as the gold standard for presence or absence of stones. Data were collected from 119 patients on the resource use related to each investigation, including hospitalisations and treatment of associated morbidity. Cost data were drawn from the literature and charges from French private clinics. The analysis models the potential cost and diagnostic success of three different strategies for the use of the investigations. The use of EUS as the first-line investigation was more costly but might leave fewer undiagnosed stones. No formal cost-effectiveness analysis was carried out, although the relevant data seemed to be available. In the interpretation of the results it was recognised that younger patients had been excluded from the study because endoscopic sphincterotomy was not always an appropriate procedure for them.

The immediate conclusion from review of the literature on economic aspects is that there are no good, comprehensive economic studies of EUS from which conclusions about its cost-effectiveness in the staging of gastrointestinal cancer can be drawn. However, some useful information on the direct costs of the procedure can be located and, as discussed in sections 3.2 and 3.3, some data on staging and therapeutic impact can be found. Data on patient outcomes that are suitable for more sophisticated types of economic analysis, such as cost–utility analysis, are not available. Given the disparate sources of this information, the only feasible approach to the economic evaluation of EUS at present is decision modelling. This approach is discussed further in chapter 9

8.6 Changes in the knowledge base

Primary studies were included in this review only if they were fully published before January 1997. Although the full search and review protocol has not been performed on studies published in 1997, we are aware of a number recent publications with findings of relevance to this review. The strategy of contacting authors, academic centres and manufacturers described in section 3.1.1 resulted in information about 19 studies that had been submitted for publication, were in press or were not then complete. Of nine incomplete studies, four are particularly notable: a cost–benefit analysis; a therapeutic impact study; a staging impact study comparing EUS and MRI; and a study of FNA.

Figure 8.2 shows the number of staging performance articles plotted by year. The shaded bars represent the number passing all the inclusion

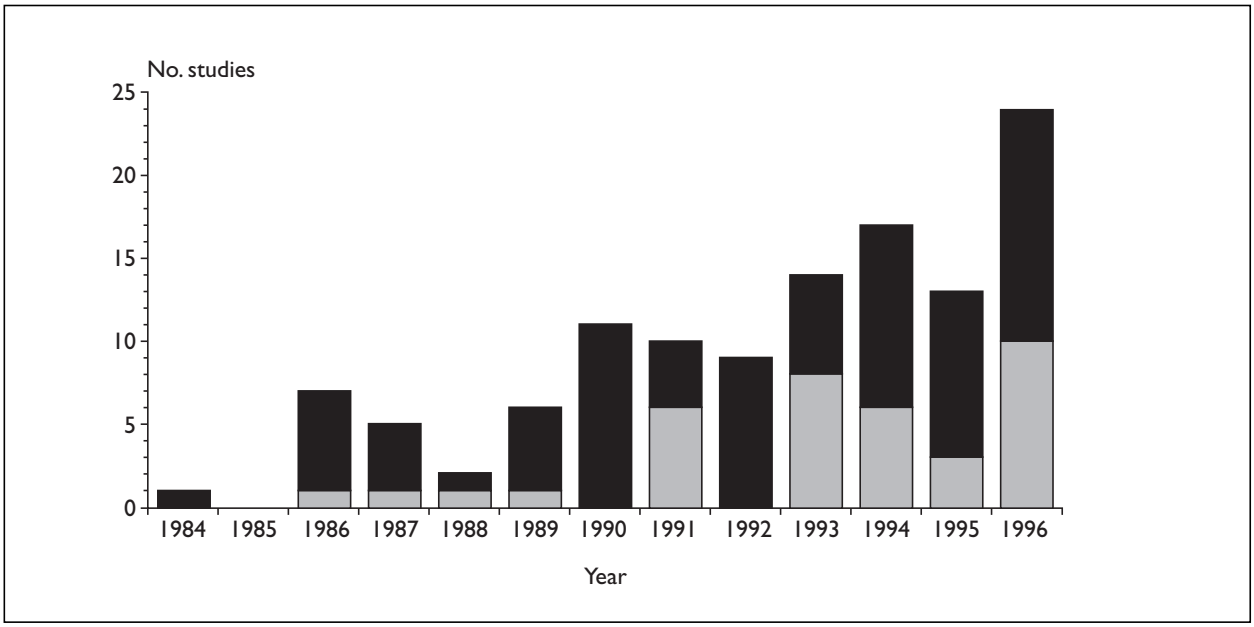


FIGURE 8.2 *Number of studies included in and excluded from this review, plotted by year of publication; includes oesophageal, gastric and cardia tumour-staging studies (□, no. passing all inclusion criteria; ■, no. excluded from review)*

criteria of this review, and the solid bars those excluded from the review by using the criteria of *Table 3.4*. Both bars together represent the total number of studies addressing the staging performance of oesophageal or gastric cancer. It can be seen that generally there has been a steady increase in the number of published studies. There is no indication of reaching a plateau or of a sudden increase or decrease in publications. From the trend, fewer than 50 articles per annum will need to be retrieved and undergo the assessment process for the next 3 years. With reference to the inclusion criteria set by this review, it can be seen that the number of studies satisfying the criteria peaked in 1993 and then fell in 1994 and 1995, before recovering in 1996. Whether this recent improvement in quality is set to continue cannot be determined from present data. The changes in the TNM classification scheme from 1997[3] will need to be recognised in updated reviews.

There is likely to be a small increase in the numbers of published studies at the other levels of the Fineberg framework, but the current numbers are so small that it is probable that reading these articles will not be an onerous task.

Chapter 9

Implications of the review

9.0 Review methodology

MEDLINE and BIDS provided the bulk of the primary studies included in our review. Much time and effort was spent seeking out further articles. Their impact on the final result has not been evaluated. As timeliness is so important in the systematic reviewing field, a methodological study to assess the impact of excluding studies that are not on the major databases would be of interest to determine whether reviews following systematic principles, but relying on the major databases alone, give valid results. Sufficient systematic reviews have now been carried out under the HTA programme for a retrospective analysis across several subjects to be feasible. It was particularly noteworthy that no further health economics studies were identified across two medical imaging topics,[17] suggesting that relatively speedy health economics reviews may be feasible.

The assessment of study validity was problematic. The difficulties we experienced in trying to develop an objective validity scoring system were vindicated by the regression analysis, which investigated the impact of study design features upon the result it gave for staging performance. At the significance level equivalent to $p < 0.05$ after the Bonferroni correction, none of the features was significant. This was partly related to widespread omissions in the reports, which made it hard to determine exactly how a study had been performed and what its strengths and weaknesses were. It would be sensible for more discussion about assessing study validity to take place between those undertaking reviews in areas where RCTs do not exist. This would prevent numerous different methodologies being introduced and might result in a single robust one being agreed. There is even scope for comparative trials to be performed by using results from existing reviews. This might be coordinated by the Cochrane Methods Working Group on Systematic Reviews of Screening and Diagnostic Tests.[22]

In this review it was possible to perform some of the alternative techniques for data synthesis suggested in the literature.[25,26] The EWLS method was compared with the RR method; the findings suggested that the EWLS method was preferable. The experiments on the exclusion of points on an ROC scatterplot lying outside a range of clinical utility defined as both sensitivity and specificity greater than 50% were inconclusive. This would be a fruitful area for further research, probably by using data from a systematic review that included rather more studies than this one.

There is a vast literature on publication bias,[13,40] which may lead a novice reviewer to believe that this is a significant problem in all subject areas. Publication bias will have most impact where the effect sought or measured by the primary studies is small, especially when some studies have positive and some negative findings. In this review of staging performance, results were much more clear cut and it is hard to predict what performance threshold would represent negative findings likely to remain unpublished. Efforts were made to find unpublished studies that may have fallen victim to publication bias but none was found. In the medical imaging field, we hypothesise that publication bias does not occur for high-quality, well-designed studies of techniques that have high sensitivity and specificity. The bias may be more prevalent for incremental imaging developments and for poorly designed studies giving equivocal results, which have been excluded from this review. This is another area where further research would help to guide future reviews, answering the question of whether publication bias is likely to be significant in this topic area.

For this review, the importance of having a multidisciplinary review panel cannot be overemphasised. The subtleties of the staging process meant that the inclusion of at least one clinical expert was vital. It would have been impossible for lay reviewers, no matter how skilled in the review process, to formulate the questions to address and extract the correct information from the articles.

9.1 EUS in gastro-oesophageal cancer – staging performance

The results of the review indicate that EUS may be recommended, for its performance if not yet its cost-effectiveness, for tumour T staging in the oesophagus and stomach. Further research is required into the use of this technology at the cardia; this is particularly important because the

incidence of cardia tumours is increasing. Current evidence does not support the use of EUS for metastasis staging. There is a gap in the evidence about the use of EUS together with another modality, such as spiral CT, so that the complementary strengths and weaknesses of the technologies may be exploited.

There is a lack of consistent information about the proportion of patients undergoing EUS who will have an incomplete study owing to non-traversability, and whether an incomplete study will influence the result. This is an area where further primary and secondary research is indicated.

There is potential for a review of EUS in pancreaticobiliary cancer.

9.2 EUS in gastro-oesophageal cancer – staging impact

There is little direct evidence available in the literature about the staging impact of EUS. The only comparative studies published relate to incremental CT; there is none regarding spiral CT, MRI or positron emission tomography. There is a need for well-designed studies to be performed (according to the comments in chapter 1), using the optimal protocols for both EUS and spiral CT/MRI/ positron emission tomography. Any call for proposals should specify that imaging equipment and planned protocols must be described. Determination should be made at the refereeing stage that these represent the best achievable for each modality, so that studies do not unduly favour a championed technology. Full pathological comparison and long-term QoL assessment must be included, so that results are useful at higher levels of the Fineberg hierarchy. As already indicated, the role of EUS is likely to be as a complementary technique to the comparator, and any further study must have a comparison not only of the two technologies used separately but an investigation of their capabilities when used together.

There is insufficient evidence in the literature for a recommendation to be made to potential purchasers about the relative merits of dedicated EUS systems and special probes that may be used with existing ultrasound equipment. A well-designed (according to the comments in chapter 1) comparative study in this area is needed; it must also include health economics analysis.

EUS-guided FNA has been insufficiently investigated. A study that concentrates on the lymph nodes is required, which defines clearly its pathological requirements to avoid problems that arise in relation to biopsies. Dye would be used to ensure that the same lymph nodes are investigated by both imaging and confirmatory techniques. Morbidity and mortality should be assessed.

A study to determine the value of miniprobes prior to endoscopic mucosal resection is recommended.

The evidence on reproducibility and the learning curve is sparse, as has been noted in previous publications on training.[41] To provide the professional bodies with sufficient information on which to base recommendations there are two main requirements: first, further retrospective studies to confirm the observations on a single observer presented by Fockens *et al.*;[47] and, secondly, studies at the diagnostic impact and higher levels of the Fineberg framework to determine the consequences of staging errors for lymph nodes in different areas.

9.3 EUS in any clinical application – therapeutic impact

There is a lack of evidence at the therapeutic impact level of the Fineberg framework and a need for studies in this area. Sufficiently extended follow-up to observe changes in management must be included and, to avoid the problems described in chapter 1, which occur if only the therapy that was actually given is recorded, a design setting out intentions to treat independently for the compared technologies is recommended. Any data at this level are also of value in full economic assessments, as discussed in section 9.5.

9.4 EUS in any clinical application – patient outcome

Again, there is a lack of evidence at the patient outcome level of the Fineberg framework and a need for new studies. Sufficiently extended follow-up to observe outcome must be included, using the QoL indicators described in chapter 1. Any data at this level are also of value in full economic assessments, as discussed in section 9.5.

9.5 EUS in any clinical application – health economics

It was noted in the previous chapter that the only feasible approach to economic evaluation of EUS at present is decision modelling. This approach is familiar to economists and is being used increas-

ingly in the clinical field. The principal advantage of modelling is that the analysis can be designed to address a specific question and can draw the best available data from multiple sources. The main disadvantage is the need to make assumptions where suitable data do not exist. Although the effect of assumptions can be tested through sensitivity analysis, the danger of bias being introduced is real. Good reviews of the use of modelling in economic evaluation in health care can be found in Sheldon[42] and Buxton *et al.*[43] The general agreement is that modelling is useful in generating and selecting hypotheses prior to the design of clinical trials and in extending the application of clinical trial data to different patient groups, different care settings, and different periods. This last point is particularly important in the evaluation of diagnostic techniques as the time between investigation and final patient outcome is often quite long and few trials follow-up for that length of time.

Decision modelling is useful only if the data sources are reliable. Considerable improvement is necessary in the quality of data collected in different types of study (e.g. cost, impact and outcome) before good evaluations of EUS could be carried out using this approach. It would be unrealistic to expect the conduct of evaluations of diagnostic technologies to be transformed overnight to meet all the criteria specified in Drummond *et al.*[30] or the *British Medical Journal*.[10] This should be regarded as a long-term objective. In the meantime, practical steps can be taken to improve the quality of information without a major additional burden on investigators.

For example, in estimating costs, the application of some simple principles could bring about a great improvement. The impact of the use of a diagnostic procedure on the use of healthcare resources should be measured comprehensively, including the impact on the use of other tests and the treatment ultimately given. Many diagnostic technologies involve significant capital investment; it is vital to take account of this when distinguishing between marginal and average costs over different times. Too much reliance should not be placed on hospital charges as a source of unit costs. As the period of studies is extended, the importance of discounting costs must be understood.

Although the use of patient-based outcome measures in clinical studies is increasing, it is by no means universal, so success in this endeavour is not guaranteed. It is therefore important that as many new studies as possible in the diagnostic field should include outcome measures that can ultimately be used in economic evaluations.

Chapter 10

Dissemination and further research

The important target audiences for dissemination of the results of this review are the purchasing decision makers in the NHS and potential users of the technology (radiologists and surgeons). In the absence of sound economic data, no firm recommendations can be made to purchasers. The results at the staging performance level are, however, encouraging, especially for T staging of gastric and oesophageal cancer. Because EUS is less suited to M staging and because studies may not be completed due to non-traversability, it should not be used without a complementary technique such as CT. There is currently insufficient evidence for recommendations to be made about the use of dedicated equipment, the role of mini-probes, EUS-guided FNA or the minimum number of examinations required for users to achieve competence. Recommendations for research in these areas have been made in chapter 9.

In the studies reported in the literature reviewed a number of design faults were particularly common. The new studies that we have recommended in chapter 9 for staging performance and impact should be designed to:

- use just one reference standard, ideally the recognised gold standard
- avoid verification bias by ensuring the one reference test is applied to all subjects
- ensure that observers are blinded to the results of other studies and particularly to the reference result
- include randomisation
- publish sufficient data for completion of 2×2 contingency tables
- use published recommendations on sample size calculation to ensure that enough subjects for each anatomical area are included for statistical validity
- comment on operator dependence/learning curves
- publish study design information to allow proper assessment of study quality.

In summary, the lessons learned from the health economics section of this review suggest the following strategy for economic evaluation in the staging field:

- clarify staging accuracy from good quality studies
- establish staging and therapeutic impact from such studies where possible
- encourage new studies specifically to measure staging impact, therapeutic impact and patient outcome
- encourage the use of patient-based outcome measures
- estimate costs from good-quality studies
- use decision-modelling techniques to combine outcome and cost data from the different sources.

The ultimate target audience is the entire medical imaging community, including those who perform studies, write articles, referee articles and edit journals. A key point noted in this review was the poor quality of the descriptions of the studies reported in the literature. It was not possible properly to assess study design because the pertinent information was missing. In the interests of both facilitating secondary research and improving the quality of primary studies, the medical imaging community must be made aware of the importance of not only designing a study well but also of reporting the features of that design in a comprehensive manner.

Acknowledgements

The authors thank the following people for their invaluable contributions to the review.

External panel member:
Professor Ian Isherwood

Statistical support:
Sub-Unit for Medical Statistics (SUMS), University of Leeds

Literature searching assistance:
Medical and Dental Library Staff, University of Leeds
Ms K Saunders, British Institute of Radiology

Secretarial support:
Mrs J Pemberton, University of Leeds
Mr PE Lines

Checklist testers:
Dr TL Freeman, University of Leeds
Mr DJ Robinson, University of Leeds
Dr SF Tanner, University of Leeds
Mr JG Truscott, University of Leeds

Respondents to information requests:
Mr M Andrews, Philips Medical Systems
Mr M Haydon, Philips Medical Systems
Mr J Stoddart, Siemens Medical Engineering
Ms W Wetherfield, Toshiba Medical Systems
Dr P Vilmann, Gentofte University Hospital, Denmark
Dr Allescher, Technical University of Munich, Germany
Dr Murata, Tokyo Womens Medical College, Japan
Dr Brugge, Massachusetts General Hospital, USA
Dr JD Scheel-Hincke, Odense, Denmark
Dr Rumberger, Mayo Clinic Foundation, USA
Dr Tierney, University of Michigan Medical Center, USA
Dr Levy, Hospital Henri Mondor, France
Dr Cellier, Laennec Hospital, Paris, France
Dr Souquet, Hospital E Herriot, France
Dr Napolean, Clinique St-Jean, Lyon, France
Dr Pollack, Case Western Reserve University, USA
Dr Wiersema, St Vincent Hospitals and Health Care Center, USA
Dr N Harada, Chiba University School of Medicine, Japan
Dr Binmoeller, University Hospital Eppendorf, Germany
Dr Erickson, Texas A&M, USA
Dr Pohl, University of Cologne, Germany
Dr H Gerdes, Memorial Sloan-Kettering Cancer Center, USA
Dr Barmeir, Chaim Sheba Medical Center, Israel
Dr Palazzo, Beujon Hospital, France
Dr H Yoshikane, Nagoya University School of Medicine, Japan
Dr P Fockens, Academic Medical Centre, Amsterdam, The Netherlands
Dr R Castro, Hôpital Ambroise Pare, France
Dr BD Greenwald, University of MD Medical School, USA
Dr S-O Molin, Sahlgren's University Hospital, Sweden
Dr AM Kassem, Technical University of Munich, Germany
Professor SA McClave, University of Louisville, KY, USA
Dr M Hiele, University Hospital Gasthuisberg, Leuven, Belgium
Dr J Vickers, Bristol Royal Infirmary, UK
Dr CJ Lightdale, Columbia Presbyterian Medical Center, New York, USA

This work was carried out with the financial support of The Secretary of State for Health under the NHS Health Technology Assessment Programme, project 94/44/03. The views and opinions expressed do not necessarily reflect those of The Secretary of State for Health. In part, this work was undertaken by the Leeds Hospitals NHS Trust, who received funding from the NHS Executive; the views expressed in this publication are those of the authors and not necessarily those of the NHS Executive.

References

The references are indicated in the text by superscripted numerals. They are divided into three sections:

- references cited in the review: superscripts with **no brackets**
- references included in the review: superscripts with **square brackets**
- references excluded from the review: superscripts with **round brackets**.

This method of citation is maintained throughout.

Cited references

1. Office of Population Censuses and Surveys. Mortality statistics (DH2). London: HMSO, 1997:22.
2. Hermanek P, Sobin LH, editors. TNM classification of malignant tumours. 4th ed. Heidelberg: Springer-Verlag, 1987.
3. Sobin LH, Wittekind C, editors. TNM classification of malignant tumours. 5th ed. New York: Springer-Verlag, 1997.
4. Fineberg HV, Bauman R, Sosman M. Computerized cranial tomography. Effect on diagnostic and therapeutic plans. *JAMA* 1977;**238**:224–7.
5. Mackenzie R, Dixon AK. Measuring the effects of imaging: an evaluative framework. *Clin Radiol* 1995;**50**:513–8.
6. Kelly S, Berry E, Roderick P, Harris KM, Cullingworth J, Gathercole L, *et al.* The identification of bias in studies of the diagnostic performance of imaging modalities. *Br J Radiol* 1997;**70**:1028–35.
7. Spilker B. Quality of life and pharmacoeconomics in clinical trials. 2nd ed. Philadelphia, PA: Lippincott-Raven; 1996.
8. Guyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook DJ. Users' guides to the medical literature: XII. How to use articles about health-related quality of life. Evidence-Based Medicine Working Group. *JAMA* 1997;**277**:1232–7.
9. Fries J, Singh G. The hierarchy of patient outcomes. In: Spilker B, editor. Quality of life and pharmacoeconomics in clinical trials. 2nd ed. Philiadelphia, PA: Lippincott-Raven, 1996:33–40.
10. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the *BMJ*. The BMJ Economic Evaluation Working Party. *BMJ* 1996;**313**:275–83.
11. Williams A. Economics of coronary artery bypass grafting. *BMJ (Clin Res)* 1994;**1985**:326–9.
12. Bryan S. Economic evaluation of diagnostic technologies: a rule for conjoint analysis? *Soc Sci Health* 1997;**3**:209–11.
13. NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness (CRD Report 4). York: University of York, 1996.
14. The British Library: Inside Information Plus. 1998: http://www.bl.uk/online/inside
15. Online Computer Library Center. 1998: http://www.oclc.org
16. American Endosonography Club. 1998: http://www.duke.edu/eus.html
17. Berry E, Kelly S, Hutton J, Harris KM, Roderick P, Boyce JC, *et al.* A systematic literature review of spiral and electron beam CT. *Health Technol Assess* 1997: In preparation.
18. NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness (CRD Report 4). York: University of York, 1996:33.
19. Mulrow CD, Linn WD, Gaul MK, Pugh JA. Assessing quality of a diagnostic test evaluation. *J Gen Intern Med* 1989;**4**:288–95.
20. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature: III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;**271**:703–7.
21. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature: III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994;**271**:389–91.
22. Cochrane Methods Working Group on Systematic Reviews of Screening and Diagnostic Tests. Recommended methods, updated 6 June 1996. http://som.flinders.edu.au/FUSA/COCHRANE/cochrane/sadtdoc1.htm

23. Berry E, Kelly S, Harris KM, Cullingworth J, Gathercole L, Hutton J, *et al.* Development of a methodology for systematic literature reviews in diagnostic imaging. *Br J Radiol* 1997;**70**(Suppl):109.

24. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* 1989;**24**:234–45.

25. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993;**12**:1293–316.

26. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, *et al.* Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994;**120**:667–76.

27. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;**48**:119–30.

28. Williams A. The cost–benefit approach. *Br Med Bull* 1974;**30**:252–6.

29. McMaster University. How to read clinical journals: VII. To understand an economic evaluation (part B). *Can Med Assoc J* 1984;**130**:1542–9.

30. Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 1997.

31. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Med Care* 1992;**30**:231–43.

32. Udvarhelyi IS, Colditz GA, Rai A, Epstein AM. Cost-effectiveness and cost–benefit analyses in the medical literature. Are the methods being used correctly? *Ann Intern Med* 1992;**116**:238–44.

33. Hutton J, Clark M, Sanderson D. The cost-effectiveness of MRI in the DGH: a review. Report to the NHS Executive HTA Programme. London: MEDTAP, 1998: In preparation.

34. European Society of Gastrointestinal Endoscopy. 1998; http://www.eus-online.org

35. Gore SM. Statistics in practice. London: British Medical Association, 1982.

36. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;**309**:1286–91.

37. Thompson WM, Halvorsen RA, Foster WL, Williford ME, Postlethwait RW, Korobkin M. Computed tomography for staging esophageal and gastroesophageal cancer: reevaluation. *AJR Am J Roentgenol* 1983;**141**:951–8.

38. Kent DL, Larson EB. Disease, level of impact, and quality of research methods. Three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Invest Radiol* 1992;**27**:245–54.

39. Davies J, Chalmers AG, Sue-Ling HM, May J, Miller GV, Martin IG, *et al.* Spiral computer tomography and operative staging of gastric carcinoma: a comparison with histopathological staging. *Gut* 1997;**41**:314–9.

40. Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci* 1993;**703**:135–48.

41. Boyce HW. Training in endoscopic ultrasonography. *Gastrointest Endosc* 1996;**43**:S12–15.

42. Sheldon TA. Problems of using modelling in the economic evaluation of health care. *Health Econ* 1996;**5**:1–11.

43. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, *et al.* Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997;**6**:217–27.

References of studies included

Staging performance

[1] Akahoshi K, Misawa T, Fujishima H, Chijiiwa Y, Maruoka A, Ohkubo A, *et al.* Preoperative evaluation of gastric cancer by endoscopic ultrasound. *Gut* 1991;**32**:479–82.

[2] Altorki NK, Snady H, Skinner DB. Endosonography for cancer of the esophagus and cardia: is it worthwhile? *Dis Esophagus* 1996;**9**:198–201.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] Catalano MF, Sivak MV, Rice T, Gragg LA, Van Dam J. Endosonographic features predictive of lymph node metastasis. *Gastrointest Endosc* 1994;**40**:442–6.

[8] Chang KJ, Katz KD, Durbin TE, Erickson RA, Butler JA, Lin F, *et al.* Endoscopic ultrasound-guided fine-needle aspiration. *Gastrointest Endosc* 1994;**40**:694–9.

[9] Dittler HJ, Siewart JR. Role of endoscopic ultrasonography in esophageal carcinoma. *Endoscopy* 1993;**25**:156–61.

[10] Dittler HJ, Siewart JR. Role of endoscopic ultrasonography in gastric carcinoma. *Endoscopy* 1993;**25**:162–6.

[11] François E, Peroux J-L, Mouroux J, Chazalle M, Hastier P, Ferrero J, *et al.* Preoperative endosonographic staging of cancer of the cardia. *Abdom Imaging* 1996;**21**:483–7.

[12] Giovannini M, Seitz J-F, Monges G, Perrier H, Rabbia I. Fine-needle aspiration cytology guided by endoscopic ultrasonography: results in 141 patients. *Endoscopy* 1995;**27**:171–7.

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

[14] 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.

[15] Harada N, Kouzu Y, Arima M, Isono K. Endoscopic ultrasound-guided histologic needle biopsy: preliminary results using a newly developed endoscopic ultrasound transducer. *Gastrointest Endosc* 1996;**44**:327–30.

[16] 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.

[17] Heintz A, Höhne U, Schweden F, Junginger T. Preoperative detection of intrathoracic tumor spread of esophageal cancer: endosonography versus computed tomography. *Surg Endosc* 1991;**5**:75–8.

[18] 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.

[19] Hünerbein M, Dohmoto M, Rau B, Schlag PM. Endosonography and endosonography-guided biopsy of upper-GI-tract tumours using a curved-array echoendoscope. *Surg Endosc* 1996;**10**:1205–9.

[20] Manzoni G. Endosonography and CT in the evaluation of tumour invasion. *Recent Adv Dis Esophagus* 1993:532–9.

[21] 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.

[22] Murata Y, Suzuki S, Hashimoto H. Endoscopic ultrasonography of the upper gastrointestinal tract. *Surg Endosc* 1988;**2**:180–3.

[23] Murata Y, Hayashi K, Kobayashi A, Yoshida K, Nagasako K, Ide H, *et al.* Pre-operative staging of oesophageal carcinoma by ultrasound. *Asian J Surg* 1993;**17**:200–7.

[24] Murata Y, Suzuki S, Ohta M, Mitsunaga A, Hayashi K, Yoshida K, *et al.* Small ultrasonic probes for determination of the depth of superficial esophageal cancer. *Gastrointest Endosc* 1996;**44**:23–8.

[25] Perng D-S, Jan C-M, Wang W-M, 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.

[26] Peters JH, Hoeft SF, Heimbucher J, Bremner RM, De Meester TR, Bremner GG, *et al.* Selection of patients for curative or palliative resection of esophageal cancer based on preoperative endoscopic ultrasonography. *Arch Surg* 1994;**129**:534–9.

[27] 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.

[28] Shimizu S, Tada M, Kawai K. Endoscopic ultrasonography for early gastric cancer. *Endoscopy* 1994;**26**:767–8.

[29] Takemoto T, Ito T, Aibe T, Okita K. Endoscopic ultrasonography in the diagnosis of esophageal carcinoma, with particular regard to staging it for operability. *Endoscopy* 1986;**18**(Suppl 3):22–5.

[30] Tio TL, Schouwink MH, Cikot RJLM, Tytgat GNJ. 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.

[31] Vilmann P. Endoscopic ultrasonography-guided fine-needle apiration biopsy of lymph nodes. *Gastrointest Endosc* 1996;**43**:S24–9.

[32] Vilmann P, Hancke S, Henriksen FW, Jacobsen GK. Endoscopic ultrasonography-guided fine-needle aspiration biopsy of lesions in the upper gastrointestinal tract. *Gastrointest Endosc* 1995;**41**:230–5.

[33] Wiersema MJ, Wiersema LM, Khusro Q, Cramer HM, Tao L-C. Combined endosonography and fine-needle aspiration cytology in the evaluation of gastrointestinal lesions. *Gastrointest Endosc* 1994;**40**:199–206.

[34] Yanai H, Tada M, Karita M, Okita K. Diagnostic utility of 20-megahertz linear endoscopic ultrasonography in early gastric cancer. *Gastrointest Endosc* 1996;**44**:29–33.

[35] Ziegler K, Sanft C, Friedrich M, Stein H, Häring R, Riecken EO. Evaluation of endosonography in TN staging of oesophageal cancer. *Gut* 1991;**32**:16–20.

[36] 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.

Staging impact, therapeutic impact and patient outcome

[37] Fok M, Cheng SWK, Wong J. Endosonography in patient selection for surgical treatment of esophageal carcinoma. *World J Surg* 1992;**16**:1098–103.

[38] Jafri IH, Saltzman JR, Colby JM, Krims PE. Evaluation of the clinical impact of endoscopic ultrasonography in gastrointestinal disease. *Gastrointest Endosc* 1996;**44**:367–70.

[39] Nickl NJ, Bhutani MS, Catalano M, Hoffman B, Hawes R, Chak A, *et al.* Clinical implications of endoscopic ultrasound: the American Endosonography Club Study. *Gastrointest Endosc* 1996;**44**:371–7.

[40] Ramirez JM, Mortensen NJM, Takeuchi N, Smilgin Humphreys MM. Endoluminal ultrasonography in the follow-up of patients with rectal cancer. *Br J Surg* 1994;**81**:692–4.

[41] Setti CP, Kamm MA, Nicholls RJ. Long-term results of postanal repair for neurogenic faecal incontinence. *Br J Surg* 1994;**81**:140–4.

[42] Snady H, Cooperman A, Siegel J. Endoscopic ultrasonography compared with computed tomography with ERCP in patients with obstructive jaundice or small peri-pancreatic mass. *Gastrointest Endosc* 1992;**38**:27–34.

Economics

[43] Allgayer H. Cost-effectiveness of endoscopic ultrasonography in submucosal tumors. *Gastrointest Endosc Clin North Am* 1995;**5**:625–9.

[44] Prat F, Amouyal G, Amouyal P, Pelletier G, Fritsch J, Choury AD, *et al.* Prospective controlled study of endoscopic ultrasonography and endoscopic retrograde cholangiography in patients with suspected common-bileduct lithiasis. *Lancet* 1996;**347**:75–9.

Reproducibility and learning curve

[45] Burtin P, Napolean B, Palazzp L, Roseau G, Souquet JC, Cales P, *et al.* Interobserver agreement in endoscopic ultrasonography staging of esophageal and cardia cancer. *Gastrointest Endosc* 1996;**43**:20–4.

[46] Catalano MF, Sivak MV, Bedford RA, Falk GW, van Stolk R, Presa F, *et al.* Observer variation and reproducibility of endoscopic ultrasound. *Gastrointest Endosc* 1995;**41**:115–20.

[47] Fockens P, Van den Brande JHM, van Dulleman HM, van Lanschot JJB, Tytgat GNJ. Endosonographic T-staging of esophageal carcinoma: a learning curve. *Gastrointest Endosc* 1996;**44**:58–62.

References of studies excluded

Staging performance

(1) Aibe T, Ito T, Yoshida T, Noguchi T, Ohtani T, Fuji T, *et al.* Endoscopic ultrasonography of lymph nodes surrounding the upper GI tract. *Scand J Gastroenterol Suppl* 1986;**123**:164–9.

(2) Aibe T, Fujimura H, Yanai H, Okita K, Takemoto T. Endosonographic diagnosis of metastatic lymph nodes in gastric carcinoma. *Endoscopy* 1992;**24**:315–9.

(3) 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.

(4) Akahoshi K, Chijiiwa Y, Tanaka M, Harada N, Nawata H. Endosonography probe-guided endoscopic mucosal resection of gastric neoplasms. *Gastrointest Endosc* 1995;**42**:248–52.

(5) Andriulli A, Recchia S, De Angelis C, Mazzucco D, Berti E, Arrigoni A, *et al.* Endoscopic ultrasonographic evaluation of patients with biopsy negative gastric linitis plastica. *Gastrointest Endosc* 1990;**36**:611–15.

(6) Asaki S, Nakayama Y, Ohara M, Hirasawa Y, Kanazawa N, Ota K. Comparison of the efficacy of endoscopic ultrasonography and submucosography in diagnosing the depth of gastric cancer invasion. *Tohoku J Exp Med* 1989;**159**:227–35.

(7) Bandoh T, Isoyama T, Toyoshima H. Submucosal tumors of the stomach – a study of 100 operative cases. *Surgery* 1993;**113**:498–506.

(8) Binmoeller KF, Seifert H, Soehendra N. Endoscopic ultrasonography-guided fine-needle aspiration biopsy of lymph nodes. *Endoscopy* 1994;**26**:780–3.

(9) Boku N, Ohtsu A, Fujii T, Koba I, Hosokawa K, Oda Y, *et al.* 'Para-lesional' saline injection method for assessment of small gastric cancers by endoscopic ultrasonography. *Dig Endosc* 1996;**8**:122–6.

(10) Bolondi L, Casanova P, Caletti GC, Grigioni W, Zani L, Barbara L. Primary gastric lymphoma versus gastric carcinoma: endoscopic US evaluation. *Radiology* 1987;**165**:821–6.

(11) Catalano MF, Sivak MVJ, Van Stolk R, Zuccaro GJ, Rice TW. Initial evaluation of a new-generation endoscopic ultrasound system. *Gastrointest Endosc* 1994;**40**:356–9.

(12) Catalano MF, Van Dam J, Sivak MVJ. Malignant esophageal strictures: staging accuracy of endoscopic ultrasonography. *Gastrointest Endosc* 1995;**41**:535–9.

(13) Chak A, Canto M, Gerdes H, Lightdale CJ, Hawes RH, Wiersema MJ, *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.

(14) Chandawarkar RY, Kakegawa T, Fujita H, Yamana H, Hayabuthi N. Comparative analysis of imaging modalities in the preoperative assessment of nodal metastasis in esophageal cancer. *J Surg Oncol* 1996;**61**:214–7.

(15) Chandawarkar RY, Kakegawa T, Fujita H, Yamana H, Toh Y, Fujitoh H. Endosonography for preoperative staging of specific nodal groups associated with esophageal cancer. *World J Surg* 1996;**20**:700–2.

(16) Chonan A, Fujita N, Mochizuki F, Yuki T, Ishida K, Inoue S, *et al.* Diagnosis of the depth of advanced gastric cancer invasion by endoscopic ultrasonography. Effectiveness of the balloon-compression method. *Dig Endosc* 1995;**7**:220–5.

(17) Dancygier H, Classen M. How can we diagnose the depth of cancer invasion in the esophagus? *Endoscopy* 1986;**18**(Suppl 3):19–21.

(18) Dancygier H, Classen M. Endoscopic ultrasonography in esophageal diseases. *Gastrointest Endosc* 1989;**35**:220–5.

(19) Date H, Miyashita M, Sasajima K, Toba M, Yamashita K, Takubo K, *et al.* Assessment of adventitial involvement of esophageal carcinoma by endoscopic ultrasonography. *Surg Endosc* 1990;**4**:195–7.

(20) Fockens P, Van Dullemen HM, Tytgat GN. Endosonography of stenotic esophageal carcinomas: preliminary experience with an ultra-thin, balloon-fitted ultrasound probe in four patients. *Gastrointest Endosc* 1994;**40**:226–8.

(21) Francioni F, Nishihira T, Masuda M, Kitamura M, Akaishi T, Shineha R, *et al.* The significance of preoperative diagnosis of esophageal cancer using esophageal mediastinal ultrasonography. *Tohoku J Exp Med* 1987;**152**:1–14.

(22) Frank N, Grieshammer B, Zimmermann W. A new miniature ultrasonic probe for gastrointestinal scanning: feasibility and preliminary results. *Endoscopy* 1994;**26**:603–8.

(23) Frank N, Wenk A, Holzapfel P, Fuchs T. An endoscopic ultrasound miniprobe system for esophageal disorders. Experience with a new prototype. *Recent Adv Dis Esophagus* 1996:145–54.

(24) Fujishima H, Misawa T, Chijiwa Y, Maruoka A, Akahoshi K, Nawata H. Scirrhous carcinoma of the stomach versus hypertrophic gastritis: findings at endoscopic US. *Radiology* 1991;**181**:197–200.

(25) Fujishima H, Chijiiwa Y, Nawata H. Short communication: detection of early scirrhous carcinoma of the stomach by endoscopic ultrasonography. *Br J Radiol* 1996;**69**:661–4.

(26) Fukuda M. Endoscopic sonography: use of the echoendoscope and echolaparoscope in the diagnosis of intra-abdominal disorders. *Ultrasound Annu* 1986:141–70.

(27) Glover JR, Sargeant IR, Bown SG, Lees WR. Non-optic endosonography in advanced carcinoma of the esophagus. *Gastrointest Endosc* 1994;**40**:194–8.

(28) Granstrom L, Stockeld D, Tisell A, Backman L, Aberg B. Endosonography in patients with cancer of the esophagus or the gastroesophageal junction. *Surg Res Commun* 1993;**14**:137–42.

(29) Grimm H, Binmoeller KF, Soehendra N. Ultrasonic esophagoprobe (prototype 1). *Gastrointest Endosc* 1992;**38**:490–3.

(30) Grimm H, Hamper K, Binmoeller KF, Soehendra N. Enlarged lymph nodes: malignant or not? *Endoscopy* 1992;**24**(Suppl 1):320–3.

(31) Grimm H. Endoscopic ultrasonography with the ultrasonic esophagoprobe. *Endoscopy* 1994;**26**:818–21.

(32) Heyder N, Lux G. Malignant lesions of the upper gastrointestinal tract. *Scand J Gastroenterol Suppl* 1986;**123**:47–51.

(33) Hoffman BJ, Bhutani MS, Sanders-Cliette A, Reed C, Silvestri G, McKnight J, *et al.* Endoscopic ultrasound (EUS) guided fine needle aspiration (FNA) for diagnosis of malignant lymph nodes. *Acta Endosc* 1995;**25**:473–4.

(34) Holden A, Mendelson R, Edmunds S. Preoperative staging of gastro-oesophageal junction carcinoma: comparison of endoscopic ultrasound and computed tomography. *Australas Radiol* 1996;**40**:206–12.

(35) Kallimanis GE, Gupta PK, al-Kawas FH, Tio TL, Benjamin SB, Bertagnolli ME, *et al.* Endoscopic ultrasound for staging esophageal cancer, with or without dilation, is clinically important and safe. *Gastrointest Endosc* 1995;**41**:540–6.

(36) Maruta S, Tsukamoto Y, Niwa Y, Goto H, Hase S, Yoshikane H, *et al.* Evaluation of upper gastrointestinal tumors with a new endoscopic ultrasound probe. *Gastrointest Endosc* 1994;**40**:603–8.

(37) 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.

(38) Melzer E, Avidan B, Heyman Z, Bar-Meir S. Accuracy of endoscopic ultrasonography for preoperative staging of esophageal malignancy. *Isr J Med Sci* 1995;**31**:119–21.

(39) Meyer HJ, Jahne J, Pichlmayr R. Strategies in the surgical treatment of gastric carcinoma. *Ann Oncol* 1994;**5**:S33–6.

(40) Mortensen MB, Pedersen SA, Hovendal CP. Preoperative assessment of resectability in gastro-esophageal carcinoma by linear array endoscopic ultrasonography. *Scand J Gastroenterol* 1994;**29**:341–5.

(41) Mortensen MB, Madsen MR, Hovendal CP. Pretherapeutic assessment of resectability in patients with upper gastrointestinal tract cancer by using a combination of endoscopic ultrasonography (EUS) and laparoscopy. *Surg Endosc* 1995;**9**:990–3.

(42) Mortensen MB, Scheelhincke JD, Madsen MR, Qvist N, Hovendal C. Combined endoscopic ultrasonography and laparoscopic ultrasonography in the pretherapeutic assessment of resectability in patients with upper gastrointestinal malignancies. *Scand J Gastroenterol* 1996;**31**:1115–19.

(43) Murata Y, Muroi M, Yoshida M, Ide H, Hanyu F. Endoscopic ultrasonography in the diagnosis of esophageal carcinoma. *Surg Endosc* 1987;**1**:11–16.

(44) Murata Y, Ohta S, Suzuki K, Mitsunaga A, Hayashi K, Yoshida K, *et al.* Pre-operative staging of esophageal cancer by ultrasonography. *Recent Adv Dis Esophagus* 1996:135–40.

(45) Murata Y, Suzuki S, Oota M, Hayashi K, Yoshida K, Eguchi R, *et al.* Preoperative staging esophageal cancer by endoscopic ultrasonography. XXX World Congress of the International College of Surgeons; 1996 Nov 25–29; Kyoto, Japan:253–7.

(46) Natsugoe S, Yoshinaka H, Moringa T, Shimada M, Hokita S, Baba M, *et al.* Assessment of tumor invasion of the distal esophagus in carcinoma of the cardia using endoscopic ultrasonography. *Endoscopy* 1996;**28**:750–5.

(47) 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.

(48) Nousbaum JB, Robaszkiewicz M, Cauvin JM, Calament G, Gouerou H. Endosonography can detect residual tumour infiltration after medical treatment of oesophageal cancer in the absence of endoscopic lesions. *Gut* 1992;**33**:1459–61.

(49) Odegaard S, Kimmey MB, Borkje B, Hausken T. Endoscopic ultrasonographic findings in benign and malignant diseases of the stomach. *Eur J Radiol* 1990;**11**:175–9.

(50) Ohashi S, Nakazawa S, Yoshino J. Endoscopic ultrasonography in the assessment of invasive gastric cancer. *Scand J Gastroenterol* 1989;**24**:1039–48.

(51) Okai T, Yamakawa O, Matsuda N, Kawakami H, Watanabe H, Satomura Y, *et al.* Analysis of gastric carcinoma growth by endoscopic ultrasonography. *Endoscopy* 1991;**23**:121–5.

(52) Pedersen BH, Vilmann P, Folke K, Jacobsen GK, Krasnik M, Milman N, *et al.* Endoscopic ultrasonography and real-time guided fine-needle aspiration biopsy of solid lesions of the mediastinum suspected of malignancy. *Chest* 1996;**110**:539–44.

(53) 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.

(54) Rosch T, Classen M. A new ultrasonic probe for endosonographic imaging of the upper GI-tract. Preliminary observations. *Endoscopy* 1990;**22**:41–6.

(55) Rosch T, Lorenz R, Zenker K, van 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.

(56) Saisho H, Sai K, Tsuyuguchi T, Yamaguchi T, Matsutani S, Ohto M. A new small probe for ultrasound imaging via conventional endoscope. *Gastrointest Endosc* 1995;**41**:141–5.

(57) Shorvon PJ, Lees WR, Frost RA, Cotton PB. Upper gastrointestinal endoscopic ultrasonography in gastroenterology. *Br J Radiol* 1987;**60**:429–38.

(58) Siewert JR, Holscher AH, Dittler HJ. Preoperative staging and risk analysis in esophageal carcinoma. *Hepatogastroenterology* 1990;**37**:382–7.

(59) Silva SA, Kouzu T, Ogino Y, Sato H. Endoscopic ultrasonography of esophageal tumors and compressions. *J Clin Ultrasound* 1988;**16**:149–57.

(60) Smith JW, Brennan MF, Botet JF, Gerdes H, Lightdale CJ. Preoperative endoscopic ultrasound can predict the risk of recurrence after operation for gastric carcinoma. *J Clin Oncol* 1993;**11**:2380–5.

(61) Strohm WD, Classen M. Staging of gastric and esophageal carcinoma by means of endoscopic ultrasonography. *Scand J Gastroenterol* 1987;**22**:17–21.

(62) Sugimachi K, Ohno S, Fujishima H, Kuwano H, Mori M, Misawa T. Endoscopic ultrasonographic detection of carcinomatous invasion and of lymph nodes in the thoracic esophagus. *Surgery* 1990;**107**:366–71.

(63) Takemoto T, Yanai H, Tada M, Aibe T, Fujimura H, Murata N, *et al.* Application of ultrasonic probes prior to endoscopic resection of early gastric cancer. *Endoscopy* 1992;**24**(Suppl 1):329–33.

(64) Tanaka M, Bandou T, Watanabe A, Sasaki H. A new technique in endoscopic ultrasonography of the upper gastrointestinal tract. *Endoscopy* 1990;**22**:221–5.

(65) Tio TL, Tytgat GN. Endoscopic ultrasonography in the assessment of intra- and transmural infiltration of tumours in the oesophagus, stomach and papilla of Vater and in the detection of extraoesophageal lesions. *Endoscopy* 1984;**16**:203–10.

(66) Tio TL, den Hartog Jager FC, Tytgat GN. The role of endoscopic ultrasonography in assessing local resectability of oesophagogastric malignancies. Accuracy, pitfalls, and predictability. *Scand J Gastroenterol Suppl* 1986;**123**:78–86.

(67) Tio TL, Coene PP, Schouwink MH, Tytgat GN. Esophagogastric carcinoma: preoperative TNM classification with endosonography. *Radiology* 1989;**173**:411–17.

(68) Tio TL, Cohen P, Coene PP, Udding J, den Hartog Jager FC, Tytgat GN. Endosonography and computed tomography of esophageal carcinoma. Preoperative classification compared to the new (1987) TNM system. *Gastroenterology* 1989;**96**:1478–86.

(69) Tio TL, Coene PP, den Hartog Jager FC, Tytgat GN. Preoperative TNM classification of esophageal carcinoma by endosonography. *Hepatogastroenterology* 1990;**37**:376–81.

(70) Tio TL, Coene PP, Luiken GJ, Tytgat GN. Endosonography in the clinical staging of esophagogastric carcinoma. *Gastrointest Endosc* 1990;**36**(2 Suppl):S2–10.

(71) Tio TL, Tytgat GN, den Hartog Jager FC. Endoscopic ultrasonography for the evaluation of smooth muscle tumors in the upper gastrointestinal tract: an experience with 42 cases. *Gastrointest Endosc* 1990;**36**:342–50.

(72) Toh Y, Baba K, Ikebe M, Adachi Y, Kuwano H, Sugimachi K. Endoscopic ultrasonography in the diagnosis of an early esophageal carcinoma. *Hepatogastroenterology* 1993;**40**:212–16.

(73) Van Dam J, Rice TW, Catalano MF, Kirby T, Sivak MVJ. High-grade malignant stricture is predictive of esophageal tumor stage. Risks of endosonographic evaluation. *Cancer* 1993;**71**:2910–17.

(74) 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. *AJR Am J Roentgenol* 1990;**155**:277–81.

(75) Vilmann P, Khattar S, Hancke S. Endoscopic ultrasound examination of the upper gastrointestinal tract using a curved-array transducer. A preliminary report. *Surg Endosc* 1991;**5**:79–82.

(76) Wegener M, Adamek RJ, Wedmann B, Pfaffenbach B. Endosonographically guided fine-needle aspiration puncture of paraesophagogastric mass lesions: preliminary results. *Endoscopy* 1994;**26**:586–91.

(77) Wiersema MJ, Hawes RH, Tao L-C, Wiersema LM, Kopecky KK, Rex DK, *et al.* Endoscopic ultrasonography as an adjunct to fine needle aspiration cytology of the upper and lower gastrointestinal tract. *Gastrointest Endosc* 1992;**38**:35–9.

(78) Yanai H, Fujimura H, Suzumi M, Matsura S, Awaya N, Noguchi T, *et al.* Delineation of the gastric muscularis mucosae and assessment of depth of invasion of early gastric cancer using a 20-megahertz endoscopic ultrasound probe. *Gastrointest Endosc* 1993;**39**:505–12.

(79) Yanai H, Yoshida T, Harada T, Matsumoto Y, Nishiaki M, Shigemitsu T, *et al.* Endoscopic ultrasonography of superficial esophageal cancers using a thin ultrasound probe system equipped with switchable radial and linear scanning modes. *Gastrointest Endosc* 1996;**44**:578–82.

(80) Yasuda K, Nakajima M, Kawai K. Endoscopic ultrasonography in the diagnosis of submucosal tumor of the upper digestive tract. *Scand J Gastroenterol Suppl* 1986;**123**:59–67.

(81) Yasuda K, Nakajima M, Kawai K. Endoscopic diagnosis and treatment of early gastric cancer. *Gastrointest Endosc Clin North Am* 1992;**2**:495–507.

(82) Yasuda K. Endoscopic ultrasonic probes and mucosectomy for early gastric carcinoma. *Gastrointest Endosc* 1996;**43**:S29–31.

(83) 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.

Staging impact, therapeutic impact and patient outcome

(84) Beynon J, Mortensen NJ, Foy DM, Channer JL, Rigby H, Virjee J. The detection and evaluation of locally recurrent rectal cancer with rectal endosonography. *Dis Colon Rectum* 1989;**32**:509–17.

(85) Cahn M, Chang K, Nguyen P, Butler J. Impact of endoscopic ultrasound with fine-needle aspiration on the surgical management of pancreatic cancer. *Am J Surg* 1996;**172**:470–2.

(86) Chak A, Canto M, Gerdes H, Lightdale CJ, Hawes RH, Wiersema MJ, *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.

(87) Dill JE, Hill S, Callis J, Berkhouse L, Evans P, Martin D, *et al.* Combined endoscopic ultrasound and stimulated biliary drainage in cholecystitis and microlithiasis – diagnoses and outcomes. *Endoscopy* 1995;**27**:424–7.

(88) Felt-Bersma RJ, Cuesta MA, Koorevaar M. Anal sphincter repair improves anorectal function and endosonographic image. A prospective clinical study. *Dis Colon Rectum* 1996;**39**:878–85.

(89) Herzog U, Boss M, Spichtin HP. Endoanal ultrasonography in the follow-up of anal carcinoma. *Surg Endosc* 1994;**8**:1186–9.

(90) Meyenberger C, Bertschinger P, Zala GF, Buchmann P. Anal sphincter defects in fecal incontinence: correlation between endosonography and surgery. *Endoscopy* 1996;**28**:217–24.

(91) Mosnier H, Guivarc'h M, Meduri B, Fritsch J, Outters F. Endorectal sonography in the management of rectal villous tumours. *Int J Colorectal Dis* 1990;**5**:90–3.

(92) Motoo Y, Okai T, Songur Y, Watanabe H, Yamaguchi Y, Mouri I, *et al.* Endoscopic therapy for early gastric cancer. Utility of endosonography and evaluation of prognosis. *J Clin Gastroenterol* 1995;**21**:17–23.

(93) Nakao A, Harada A, Nonami T, Kishimoto W, Takeda S, Ito K, *et al.* Prognosis of cancer of the duodenal papilla of Vater in relation to clinicopathological tumor extension. *Hepatogastroenterology* 1994;**41**:73–8.

(94) Nickl NJ, Cotton PB. Clinical application of endoscopic ultrasonography. *Am J Gastroenterol* 1990;**85**:675–82.

(95) Nielsen MB, Hauge C, Pedersen JF, Christiansen J. Endosonographic evaluation of patients with anal incontinence: findings and influence on surgical management. *AJR Am J Roentgenol* 1993;**160**:771–5.

(96) Nielsen MB, Dammegaard L, Pedersen JF. Endosonographic assessment of the anal sphincter after surgical reconstruction. *Dis Colon Rectum* 1994;**37**:434–8.

(97) Solomon M, Mcleod RS, Cohen EK, Simons ME, Wilson S. Reliability and validity studies of endoluminal ultrasonography for anorectal disorders. *Dis Colon Rectum* 1994;**37**:546–51.

(98) Taal BG, den Hartog Jager FC, Burgers JM, van Heerde P, Tio TL. Primary non-Hodgkin's lymphoma of the stomach: changing aspects and therapeutic choices. *Eur J Cancer Clin Oncol* 1989;**25**:439–50.

(99) Tio LT, Blank LE, Wijers OB, den Hartog Jager FCA, Van Dijk JDP, Tytgat GN. Staging and prognosis using endosonography in patients with inoperable esophageal carcinoma treated with combined intraluminal and external irradiation. *Gastrointest Endosc* 1994;**40**:304–10.

Economics

(100) Cahn M, Chang K, Nguyen P, Butler J. Impact of endoscopic ultrasound with fine-needle aspiration on the surgical management of pancreatic cancer. *Am J Surg* 1996;**172**:470–2.

(101) Campbell DM, Behan M, Donnelly VS, Oherlihy C, O'Connell PR. Endosonographic assessment of postpartum anal-sphincter injury using a 120 degree sector scanner. *Clin Radiol* 1996;**51**:559–61.

(102) Dhiman RK, Choudhuri G, Saraswat VA, Agarwal DK, Naik SR. Role of paraoesophageal collaterals and perforating veins on outcome of endoscopic sclerotherapy for oesophageal varices: an endosonographic study. *Gut* 1996;**38**:759–64.

(103) Dobashi Y, Nakamura H. Portal hypertension: influence on management. *Gastrointest Endosc Clin North Am* 1995;**5**:667–74.

(104) Engel AF, Kamm MA, Hawley PR. Civilian and war injuries of the perineum and anal sphincters. *Br J Surg* 1994;**81**:1069–73.

(105) Etzkorn KP, DeGuzman LJ, Holderman WH, Abu-Hammour A, Schlesinger PK, Harig JM, *et al.* Endoscopic drainage of pancreatic pseudocysts: patient selection and evaluation of the outcome by endoscopic ultrasonography. *Endoscopy* 1995;**27**:329–33.

(106) Houvenaeghel G, Martino M, Resbeut M, Rosello R, Perez T, Delpero JR, *et al.* Pelvic staging of advanced and recurrent gynecologic cancers – contribution of endosonography. *Gynecol Oncol* 1994;**55**:393–400.

(107) Houvenaeghel G, Delpero JR, Rosello R, Resbeut M, Viens P, Jacquemier J, *et al.* Results of a prospective study with comparison of clinical, endosonographic, computed tomography, magnetic resonance imaging and pathologic staging of advanced gynecologic carcinoma and recurrence. *Surg Gynecol Obstet* 1993;**177**:231–6.

(108) Jafri IH, Saltzman JR, Colby JM, Krims PE. Evaluation of the clinical impact of endoscopic ultrasonography in gastrointestinal disease. *Gastrointest Endosc* 1996;**44**:367–70.

(109) Meyenberger C, Bertschinger P, Zala GF, Buchmann P. Anal sphincter defects in fecal incontinence – correlation between endosonography and surgery. *Endoscopy* 1996;**28**:217–24.

(110) Meyenberger C, Boni RAH, Bertschinger P, Zala GF, Klotz HP, Krestin GP. Endoscopic ultrasound and endorectal magnetic-resonance-imaging – a prospective, comparative study for preoperative staging and follow-up of rectal cancer. *Endoscopy* 1995;**27**:469–79.

(111) Milsom JW, Lavery IC, Stolfi VM, Czyrko C, Church JM, Oakley JR, *et al.* The expanding utility of endoluminal ultrasonography in the management of rectal cancer. *Surgery* 1992;**112**:832–41.

(112) Mortensen MB, Pedersen SA, Hovendal CP. Preoperative assessment of resectability in gastroesophageal carcinoma by linear-array endoscopic ultrasonography. *Scand J Gastroenterol* 1994;**29**:341–5.

(113) Neimark S, Jonas SK. Obstructive jaundice: diagnostic and therapeutic considerations. *Postgrad Med* 1985;**78**:127–32.

(114) Nickl NJ, Bhutani MS, Catalano M, Hoffman B, Hawes R, Chak A, *et al.* Clinical implications of endoscopic ultrasound: the American Endosonography Club Study. *Gastrointest Endosc* 1996;**44**:371–7.

(115) Nielsen MB, Dammegaard L, Pedersen JF. Endosonographic assessment of the anal sphincter after surgical reconstruction. *Dis Colon Rectum* 1994;**37**:434–8.

(116) Obara K, Kuwana T, Ishihata R, Kondo Y, Ejiri Y, Yokogi K, *et al.* Clinical effectiveness of lansoprazole in patients with gastric-ulcers – evaluation of quality of ulcer healing based on endoscopic ultrasonographic findings. *J Clin Gastroenterol* 1995;**20**:S36–9.

(117) Parente F, Petrillo M, Vago L, Porro GB. The watermelon stomach – clinical, endoscopic, endosonographic, and therapeutic aspects in 3 cases. *Endoscopy* 1995;**27**:203–6.

(118) Pelika A, Kothaj P, Winter P, Molnar P. The preoperative staging of rectal carcinoma. *Saudi Med J* 1994;**15**:351–3.

(119) Rifkin MD, Marks GJ. Transrectal US as an adjunct in the diagnosis of rectal and extrarectal tumors. *Radiology* 1985;**157**:499–502.

(120) Romano G, Esercizio L, Santangelo M, Vallone G, Santangelo ML. Impact of computed tomography vs intrarectal ultrasound on the diagnosis, resectability, and prognosis of locally recurrent rectal cancer. *Dis Colon Rectum* 1993;**36**:261–5.

(121) Schafer A, Enck P, Furst G, Kahn T, Frieling T, Lubke HJ. Anatomy of the anal sphincters – comparison of anal endosonography to magnetic resonance imaging. *Dis Colon Rectum* 1994;**37**:777–81.

(122) Setti CP, Kamm MA, Nicholls RJ. Long-term results of postanal repair for neurogenic faecal incontinence. *Br J Surg* 1994;**81**:140–4.

(123) Shim CS, Joo JH, Park CW, Kim YS, Lee JS, Lee MS, *et al.* Effectiveness of endoscopic ultrasonography in the diagnosis of choledocholithiasis prior to laparoscopic cholecystectomy. *Endoscopy* 1995;**27**:428–32.

(124) Snady H, Cooperman A, Siegel J. Endoscopic ultrasonography compared with computed tomography with ECRP in patients with obstructive jaundice or small peri-pancreatic mass. *Gastrointest Endosc* 1992;**38**:27–34.

(125) Solomon MJ, Mcleod RS, O'Connor BI, Cohen Z. Assessment of peripouch inflammation after ileoanal anastomosis using endoluminal ultrasonography. *Dis Colon Rectum* 1995;**38**:182–7.

(126) Spada I, Taylor G, McWeeny K. Endoscopic ultrasonography. *Gastroenterol Nurs* 1990;**13**:24–30.

(127) Thompson NW, Czako PF, Fritts LL, Bude R, Bansal R, Nostrant TT, *et al.* Role of endoscopic ultrasonography in the localization of insulinomas and gastrinomas. *Surgery* 1994;**116**:1131–8.

(128) Vancutsem E, Rutgeerts P. Diagnosis and treatment with endoscopy. *Curr Opin Gastroenterol* 1994;**10**:600–4.

(129) Yasuda K, Nakajima M, Yoshida S, Kiyota K, Kawai K. The diagnosis of submucosal tumors of the stomach by endoscopic ultrasonography. *Gastrointest Endosc* 1989;**35**:10–5.

Authors' publications/ presentations

Berry E, Beckmann EC. The systematic literature review: what it is and how information technology can help. *info*RAD™ exhibit at Radiology UK 96 (May 1996, Birmingham).
Abstract: *Br J Radiol* 1996;**69**(Suppl):268.
This presentation is now available on the Internet at: http://agora.leeds.ac.uk/comir/people/eberry/sysrev/sysrev.htm

Berry E, Smith MA. Systematic reviews in diagnostic imaging.
Oral presentation at Institute of Physics and Engineering in Medicine Annual Meeting (September 1996, Leeds).

Berry E, Kelly S, Harris KM, Cullingworth J, Gathercole L, Hutton J, *et al.* Development of a methodology for systematic literature reviews in diagnostic imaging.
Poster presentation at Radiology 1997: Imaging, Science and Oncology (May 1997, Birmingham).
Abstract: *Br J Radiol* 1997;**70**(Suppl):109.

Hutton J, Berry E , Smith MA, Kelly S, Harris KM, Cullingworth J, *et al.* Technology assessment in diagnostic imaging: use of systematic literature reviews.
Oral presentation at 13th International Meeting of the International Society of Technology Assessment in Health Care (May 1997, Barcelona).

Hutton J, Berry E , Smith MA, Kelly S, Harris KM, Cullingworth J, *et al.* Technology assessment in diagnostic imaging: use of systematic literature reviews.
Oral presentation at World Congress on Medical Physics and Biomedical Engineering (September 1997, Nice).

Hutton J, Kelly S, Berry E , Smith MA. Trial design for economic evaluation of diagnostic technology.
Poster presentation at 14th International Meeting of the International Society of Technology Assessment in Health Care (June 1998, Ottawa).

Kelly S, Berry E, Roderick P, Harris KM, Cullingworth J, Gathercole L, *et al.* The identification of bias in studies of the diagnostic performance of imaging modalities. *Br J Radiol* 1997;**70**:1028–35.

Kelly S, Berry E, Roderick P, Harris KM, O'Connor PJ, Hutton J, *et al.* Medical imaging: challenges associated with the assessment of study validity in systematic literature reviews.
Oral presentation at Radiology 1998 (June 1998, Birmingham).
Abstract: *Br J Radiol* 1998;**71**(Suppl):72.

O'Connor PJ, Harris KM, Kelly S, Berry E, Hutton J, Cullingworth J, *et al.* Systematic literature review of endoscopic US and gastric cancer.
Poster presentation at Radiology 1998 (June 1998, Birmingham).
Abstract: *Br J Radiol* 1998;**71**(Suppl):98.

Appendix 1

Search strategies

Details of searches for both MEDLINE and BIDS are given for all categories owing to the slight variations between the two systems; this is most noticeable for the staging performance search strategy. The search strategy for EMBASE is shown only in section A1.1 as no modification of this strategy was performed for the other categories. *Table A1.1* gives an explanation of the abbreviations and commands used.

TABLE A1.1 *Definition of commands and abbreviations*

Abbreviation or command	Definition
$	Truncation symbol for MEDLINE
/	Heading separator for MEDLINE
ab	Abstract command for MEDLINE
adj	Adjacent command for MEDLINE
exp	Explode command for MEDLINE
hw	Subject heading word command for MEDLINE
sh	MeSH command for MEDLINE
ti	Title command for MEDLINE
tw	Textword command for MEDLINE
*	Truncation symbol for BIDS
#	Wild word command for BIDS
" - "	Unlike MEDLINE, BIDS does not ignore the hyphen, therefore hyphenated words require searching and quotation marks are necessary to avoid confusion with the alternative use of the hyphen as the Boolean NOT command.
tka	Title, keyword, abstract command for BIDS
de	Subject heading for EMBASE

A1.1 Staging performance search strategies

MEDLINE search strategy for staging performance

001 (endoscop$ adj2 ultraso$).ti,ab,sh
002 endosonograph$.ti,ab,sh
003 echoendoscop$.ti,ab,sh
004 eus.tw
005 exp endoscopy, gastrointestinal/
006 esophagoscopy/
007 5 or 6
008 ultrasonography/
009 7 and 8
010 1 or 2 or 3 or 4 or 9
011 miniprobe$.tw
012 esophagoprobe$.tw
013 blind probe$.tw
014 small probe$.tw
015 high$ frequency probe$.tw
016 (mini adj3 probe$).tw
017 (miniatur$ adj3 probe$).tw
018 (mini adj2 ultraso$ adj2 transducer$).tw
019 (miniatur$ adj2 ultraso$ adj2 transducer$).tw
020 11 or 12 or 13 ... or 19
021 10 or 20

BIDS search strategy for staging performance

001 endoscop* ultraso*.tka
002 endoscop* # ultraso*.tka
003 endoscop* # # ultraso*.tka
004 "endoscopic-ultraso*".tka
005 ultraso* endoscop*.tka
006 ultraso* # endoscop*.tka
007 ultraso* # # endoscop*.tka
008 "ultrasonic-endoscop*".tka
009 endosonograph*.tka
010 endo* sonograph*.tka
011 "endoscopic-sonograph*".tka
012 echoendoscop*.tka
013 echo* endoscop*.tka
014 "echo-endoscop*".tka
015 (ultraso* and esophagoscop*).tka
016 eus.tka
017 1 or 2 or 3 ... or 15 or 16
018 miniprobe*.tka
019 esophagoprobe*.tka
020 blind probe*.tka
021 "blind-probe*".tka

022 small probe*.tka
023 “small-probe*”.tka
024 high* frequency probe*.tka
025 “high-frequency-probe*”.tka
026 “high-frequency probe*”.tka
027 mini probe*.tka
028 mini # probe*.tka
029 mini # # probe*.tka
030 “mini-probe*”.tka
031 miniatur* probe*.tka
032 miniatur* # probe*.tka
033 miniatur* # # probe*.tka
34 “miniature-probe*”.tka
035 “miniaturised-probe*”.tka
036 “miniaturized-probe*”.tka
037 mini ultraso* transducer*.tka
038 mini # ultraso* transducer*.tka
039 miniatur* ultraso* transducer*.tka
040 miniatur* # ultraso* transducer*.tka
041 ultraso* mini transducer*.tka
042 ultraso* mini # transducer.tka
043 ultraso* miniatur* transducer*.tka
044 ultraso* miniatur* # transducer.tka
045 18 or 19 or 20 ... or 44
046 17 or 45

EMBASE search strategy for all categories

001 esophagus-cancer.de
002 endoscopy.de
003 ultrasound.de
004 2 and 3
005 endoscopic-echography.de
006 4 or 5
007 1 and 6

A1.2, A1.3 and A1.4 Staging impact, therapeutic impact and patient outcome search strategy

Owing to the similarity of some of the terminology used for the three higher levels of staging impact, therapeutic impact and patient outcome, no attempt was made to search these topics individually. A comprehensive search strategy was developed and combined with the staging performance strategy of EUS using the Boolean AND command.

MEDLINE search strategy for higher levels

001 exp survival analysis/
002 survival rate/
003 exp prognosis/
004 prognos$.tw
005 surviv$.tw
006 exp “outcome and process assessment (health care)”/
007 health.hw
008 health$.tw
009 outcome.hw
010 outcome$.tw
011 effectiveness.tw
012 efficien$.tw
013 benefi$.tw
014 improve$.tw
015 succe$.tw
016 impact.tw
017 management.tw
018 quality of life.tw
019 exp quality of life/
020 QALY.tw
021 1 or 2 or 3 ... or 20

BIDS search strategy for higher levels

001 surviv*.tka
002 prognos*.tka
003 health*.tka
004 outcome*.tka
005 effectiveness.tka
006 efficien*.tka
007 benefi*.tka
008 improve*.tka
009 succe*.tka
010 impact.tka
011 management.tka
012 quality of life.tka
013 QALY.tka
014 1 or 2 or 3 ... or 13

A1.5 Economics search strategy

MEDLINE search strategy for economics

001 exp economics/
002 cost$.tw
003 economic$.tw
004 expens$.tw
005 money$.tw
006 monetary.tw
007 financ$.tw
008 dollar$.tw
009 effectiveness.tw
010 QALY.tw
011 (benefi$ and (impact or management or outcome or utility)).tw
012 (impact and (management or outcome or utility)).tw
013 (management and (outcome or utility)).tw
014 (outcome and utility).tw

015 exp treatment outcome/ and (benefi$ or impact or management or outcome or utility).tw
016 1 or 2 or 3 ... or 15

BIDS search strategy for economics

001 cost*.tka
002 economic*.tka
003 expens*.tka
004 money*.tka
005 monetary.tka
006 financ*.tka
007 dollar*.tka
008 effectiveness.tka
009 QALY.tka
010 (benefi* and (impact or management or outcome or utility)).tka
011 (impact and (management or outcome or utility)).tka
012 (management and (outcome or utility)).tka
013 (outcome and utility).tka
014 1 or 2 or 3 ... or 13

Appendix 2

Checklists for bias and factors

A2.1 Bias checklist with specific guidelines

1. Article details

1.1 Title
1.2 Main author
1.3 What question(s) is the paper addressing?
1.4 Are these questions of value to the specific aims? Yes No

Aims: Diagnostic performance
Diagnostic impact
Therapeutic impact
Patient outcome
Cost-effectiveness

Patient selection biases

A Referral bias

Questions A1–A3 provide only information. A judgement from this information is required to assess the presence or absence of the three referral biases.

A1 Is the establishment(s) where the study was undertaken stated?

[ ] Yes = The establishment(s) is stated in the text or the origin of the establishment(s) is identifiable from the authors' correspondence addresses. The establishment is the place of origin of the study, such as a university hospital or a cancer institute.

[ ] No = It is not stated and it is unclear from which author's establishment the study was conducted.

A2 Is the establishment from where the patients were referred stated?

[ ] Yes = It is clearly stated in the text; for example, referred from local general practices.

[ ] No = It is not stated.

A3 Is the access to the establishment described?

[ ] Yes = It is stated that the establishment is open access, referral based, public or private etc.

[ ] No = No information.

B Patient filtering bias

B1 Are specific eligibility criteria stated for those included/excluded?

[ ] Yes = Criteria are either reported for all those who do receive the test or those who do not, and the total number of patients referred is given as well as the number included/excluded; or it is clear that all patients referred to the centre receive the diagnostic test.

[ ] No = Criteria or numbers are not reported.

[ ] ? = Insufficient information.

B2 Is diagnostic safety bias present or evident in the eligibility criteria?

[ ] Yes = It is clear that selected patients are excluded to avoid the 'unnecessary' diagnostic test for reasons of safety or exposure.

[ ] No = It is clear that all patients are included despite safety.

[ ] ? = Insufficient information.

B3 Is co-intervention bias present?

[ ] Yes = A selected proportion of the study group received additional interventions. Such interventions include any prior surgery, treatment or tests that are likely to influence the final test performance. This is also known as 'treatment paradox'.[22]

[ ] No = It is stated that all or none of the study group received additional interventions.

[ ] ? = Insufficient information.

B4 Is co-intervention bias avoided via the eligibility criteria?
[ ] **Yes** = It is clearly stated that patients are excluded if they have had additional interventions.
[ ] **No** = It is clear that patients were included despite co-interventions.
[ ] **?** = Insufficient information.

C Patient cohort bias

Questions C1–C3 provide only information. A judgement from this information is required to assess the presence or absence of patient cohort bias.

C1 Are the study group's clinical details described?
[ ] **Yes** = Severity or chronicity of symptoms is reported, together with sex ratio, age range and mean age of both the initial referral group and those receiving the gold standard test.
[ ] **No** = Neither severity nor chronicity, or less than three of the demographic chracters, are reported; or demographics are not given for both groups.

C2 Are the study group's pathological details described?
[ ] **Yes** = Type and location of disease are reported for those receiving the gold standard.
[ ] **No** = None or only one of the above is reported.

C3 Are the study group's co-morbid details described?
[ ] **Yes** = *Any* co-morbid conditions, or absence of conditions are reported for *any* patients.
[ ] **No** = No information regarding co-morbid conditions is reported.

Biases associated with application of the gold standard

D1 Is verification bias present?
[ ] **Yes** = Not all of the patients who have received the diagnostic test go on to receive the gold standard.
[ ] **No** = All patients receive the single gold standard test or a correction is performed by the authors.
[ ] **?** = Insufficient information.

D2 Is work-up bias present?
[ ] **Yes** = The result of the diagnostic test is used to decide who receives the gold standard.
[ ] **No** = It is clear that the diagnostic test is not used to decide, or a correction is performed by the authors.
[ ] **?** = Insufficient information.

D3 Is incorporation bias present?
[ ] **Yes** = Patients receive verification via the diagnostic test under evaluation.
[ ] **No** = The diagnostic test is not used as verification.
[ ] **?** = Insufficient information.

Biases due to the measurement of results

E Disease progression bias

E1 Is disease progression bias present for the test under evaluation?
[ ] **Yes** = The time between the diagnostic test and verification with the gold standard is greater than or equal to 21 days. (The number of days considered acceptable depends on the aetiology and understanding of the condition under review.)
[ ] **No** = The time is less than 21 days.
[ ] **?** = Insufficient information.

F Withdrawal bias

F1 Are results reported for all patients who received verification?
[ ] **Yes** = Results are clearly reported for all patients who received verification with the gold standard test.
[ ] **No** = Results are missing or selective results are reported.
[ ] **?** = Insufficient information.

F2 Are there any indeterminate test results?
[ ] **Yes** = Patients are excluded or results not reported due to indeterminate test results.
[ ] **No** = All results are included irrespective of indeterminability.
[ ] **?** = Insufficient information.

F3 Are there any patients lost to follow-up?
[ ] **Yes** = Patients are excluded or results not reported owing to loss.
[ ] **No** = All patients present for verification.
[ ] **?** = Insufficient information.

G Observer variability bias

G1 Is there a single observer of the diagnostic test under evaluation?
[ ] **Yes** = All images from the test under evaluation are interpreted by one person.
[ ] **No** = More than one interpreter.
[ ] **?** = Insufficient information.

G2 If 'no' to G1, are results reported separately for each observer?
[ ] **Yes** = All results are reported independently for all observers.
[ ] **No** = Not all results are reported separately (i.e. pooled).
[ ] **?** = Insufficient information.

G3 Is any attempt made to assess interobserver variability?
[ ] **Yes** = Data are reported statistically, with the kappa statistic, or illustrated in an ROC curve for interobserver variation.
[ ] **No** = No data are provided.
[ ] **?** = Insufficient information.

G4 Are the diagnostic test results taken from a consensus decision?
[ ] **Yes** = It is clearly stated that the test results are a consensus decision.
[ ] **No** = It is clear that it was not a consensus decision.
[ ] **?** = Insufficient information.

G5 Is any attempt made to assess intraobserver variability?
[ ] **Yes** = Data are reported statistically, with the kappa statistic, or illustrated in an ROC curve for intraobserver variation.
[ ] **No** = No data are provided.
[ ] **?** = Insufficient information.

Independence of interpretation biases

H1 Is diagnostic review bias present?
[ ] **Yes** = Observers are aware of the results of the diagnostic test when interpreting the gold standard.
[ ] **No** = It is stated that observers are blinded or unaware of the diagnostic test results.
[ ] **?** = Insufficient information.

H2 Is test review bias present?
[ ] **Yes** = Observers are aware of the results of the gold standard when interpreting the diagnostic test.
[ ] **No** = It is stated that the observers are blinded or unaware of the gold standard results.
[ ] **?** = Insufficient information.

H3 Is comparator review bias present?
[ ] **Yes** = More than one diagnostic test is compared with the gold standard and observers are aware of the result of one test when interpreting the other test.
[ ] **No** = It is stated that all the diagnostic tests were read independently or blind to the other tests; or only one diagnostic test was used.
[ ] **?** = Insufficient information.

H4 Is clinical review bias present?
[ ] **Yes** = It is stated that the observers are aware of the clinical details and history of the patients.
[ ] **No** = It is stated that the observers are blinded to the clinical data.
[ ] **?** = Insufficient information.

A2.2 Factors checklist for endoscopic ultrasound performance studies

1. Article details

1.1 Title
1.2 Main author
1.3 Over what time period was the study performed? .....

2. Study cohort

A1 Is the study randomised? Yes No ?

A2 Is the study prospective? Yes No ?

A3 Is the study controlled? Yes No ?

3. Sample size

B1 What was the total number of patients referred? .....

B2 How many patients were excluded or lost?
Before receiving test .....
After receiving test .....

B3 How many true-positive patients were there in the verified group?
T1 .....
T2 .....
T3 .....
T4 .....
N0 .....
N1 .....
N2 .....
M0 .....
M1 .....
Total .....

B4 Were patients divided into subgroups? Yes No ?

1)	11) .....
2)	12) .....
3)	13) .....
4)	14) .....
5)	15) .....
6)	16) .....
7)	17) .....
8)	18) .....
9)	19) .....
10)	20) .....

4. Clinical description

C1 Number male .....

C2 Number female .....

C3 Age range .....

C4 Mean age .....

C5 **Were symptoms/diagnosis/indications described?** Yes No ?

.....
.....
.....
.....
.....
.....

5. Homogeneity of diagnostic application

D1 **Main diagnostic application(s)** .....
.....

D2 **Diagnostic application(s) subset(s)** .....
.....
.....

D3 **Diagnostic modality(ies)** .....

D4 **Main anatomical area(s)** .....

D5 **Anatomical area(s) subset(s)** .....
.....

D6 **Tumour type(s)** .....
.....
.....
.....

6. Technical quality

E1 **Model(s)** .....

E2 **Manufacturer(s)** .....

E3 **Frequencies** .....

E4 **Radial/linear/mini** .....

7. Procedural quality

F1 **Suggested operator ability** .....

F2 **Number of readers** .....

F3 **Diagnostic criteria (thresholds/scorings)** .....
.....
.....
.....
.....
.....
.....

F4 **Was a gold standard of interoperative assessment used in unresectable cases?** Yes No ?

F5 **Did stenosis cause any impassable strictures?** Yes No ?

Appendix 3

Grading study validity

Observational studies are ranked low on the Centre for Reviews and Dissemination hierarchical scale of evidence.[18] In the absence of studies at the higher levels, as is the case for EUS, the value of the available evidence needs to be assessed on its own merits. To perform this evaluation objectively, the checklists described in section 3.1.3 were used. The information gained from these completed checklists was then used to form a hierarchy of evidence within the available studies. The methodology is described in two steps, as summarised in *Table A3.1*. For step 2, three possible methods are described in which the endpoint is a score or grade assigned to each study, which reflects the responses to the questions and estimates the likely validity of the study and its place in the hierarchy.

Throughout this methodology, a high score or definition refers to a poor study and a low score or definition refers to a more valid study. The methodology developed was applied to the 27 EUS studies included in the review of staging performance, using each of the three methods described in *Table A3.1*. The results of this trial are shown later in this appendix.

Step 1: Combining questions into groups to assign high, medium or low grading

This step is described in three sections: subjects; study; and independence of interpretations.

Each section contains flow diagrams illustrating how the questions were combined together to assign a score of high, medium or low to each bias within that section. A score of high is poor and a score of low is good. All the questions were combined into 11 biases, labelled A to L (I omitted to avoid confusion); they are summarised in *Table A3.2*.

In total, 20 biases were identified but, as some of these were related, when combined they formed the 11 groups identified. In all categories the option of the question mark ('?') referred to the report not providing sufficient information to answer the question. It has been assumed that lack of information implies the presence of bias and so question marks were grouped with the negative responses to the questions.

For disease progression bias and comparator review bias, the lack of information was scored lower than when information was provided, regardless of whether this information suggested the presence or absence of bias. These were scored in this way because, for these biases, if information was provided the likely magnitude or effect of the bias could perhaps be determined.

TABLE A3.1 *Summary of methodology*

Step	Reason
1. Combining questions into groups of biases	In the checklist, many questions are related to one particular bias or a small group of biases. By combining these questions together, a grade of high, medium or low was assigned to each bias. This reflects the likelihood of the bias being present (i.e. high refers to a high likelihood of the bias being present).
2. Combining biases into overall score	Once combined into a score per bias, the next step was to combine these further into a single overall score for the study. The three methods are briefly described below.
Method 1: Ungrouped, unweighted system	After assigning a score of 3, 2 or 1 for high, medium and low respectively, the biases were simply added together.
Method 2: Grouped, unweighted system	This method grouped the biases into the three sections as illustrated in *Table A3.4*, graded each section as high, medium or low, and assigned a score of 3, 2 or 1 as before. This time, with only three possibilities, the relevance of each section was taken into consideration in the scoring scheme.
Method 3: Grouped, weighted system	The last method grouped the biases as in method 2, but applied a weight to each set of biases before combining them into an overall score. The weight applied was in three categories, 1, 2 or 3, depending on the perceived effect of the bias (i.e. a weight of 3 for a bias that is perceived to be of greatest impact).

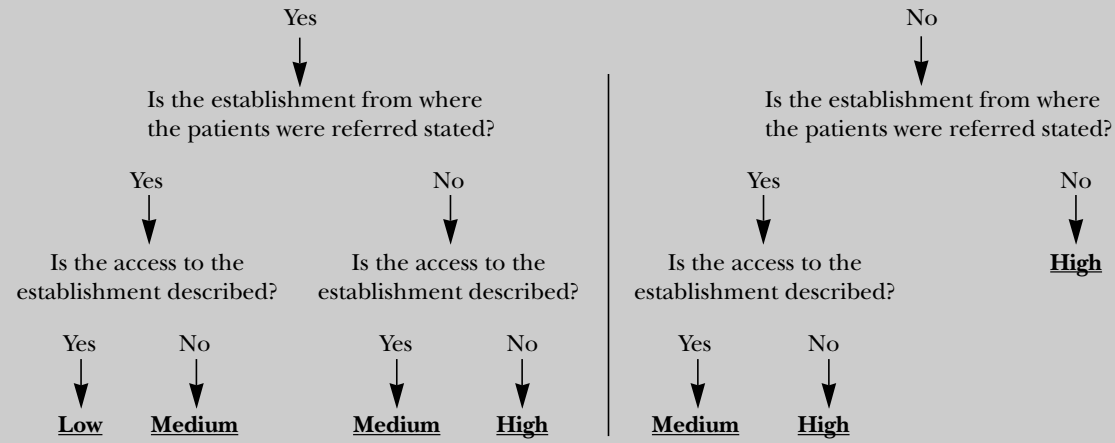
TABLE A3.2 *Summary of division of sections and biases*

Section	Subsection titles	Biases	Label
Subjects	Patient selection biases	Referral bias	A
		Patient filtering bias	B
		Patient cohort bias	C
Study	Biases associated with application of the gold standard	Verification and work-up bias	D
		Incorporation bias	E
	Biases due to the measurement of results	Disease progression bias	F
		Withdrawal bias	G
		Observer variability	H
Independence of interpretations	Biases associated with blinding between tests	Diagnostic and test review bias	J
		Comparator review bias	K
		Clinical review bias	L

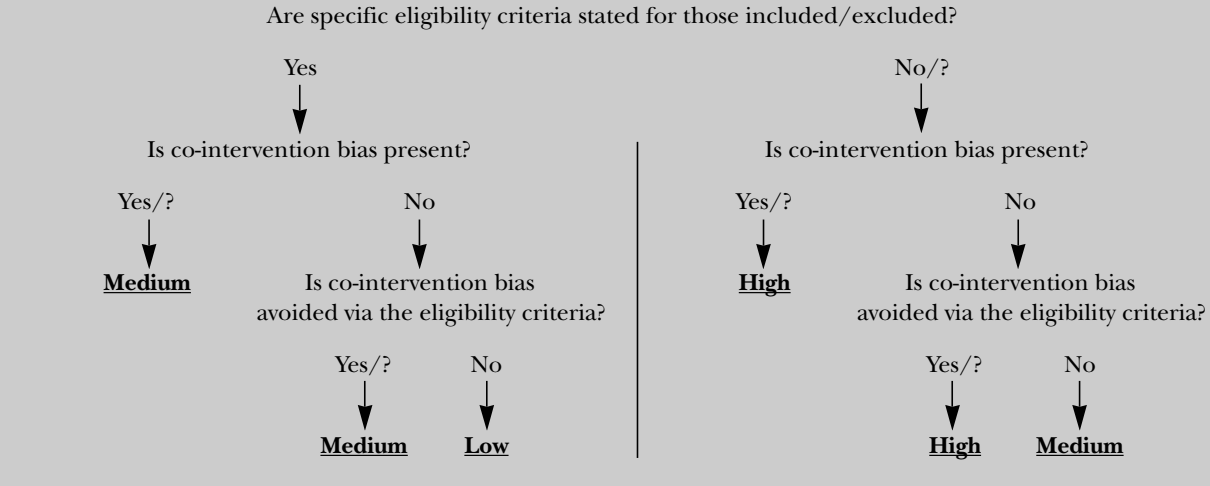
Section 1: Subjects

Patient selection biases

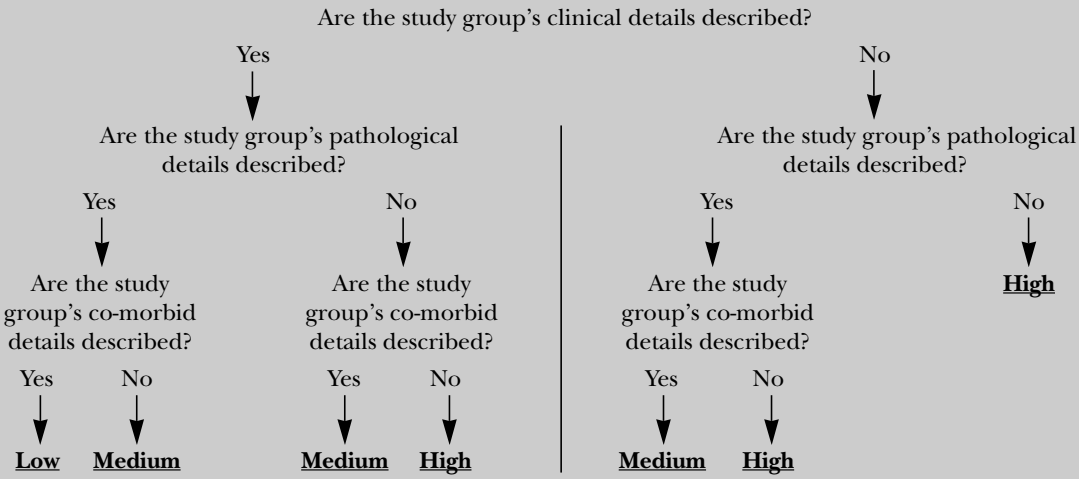
A. Referral bias



B. Patient filtering bias



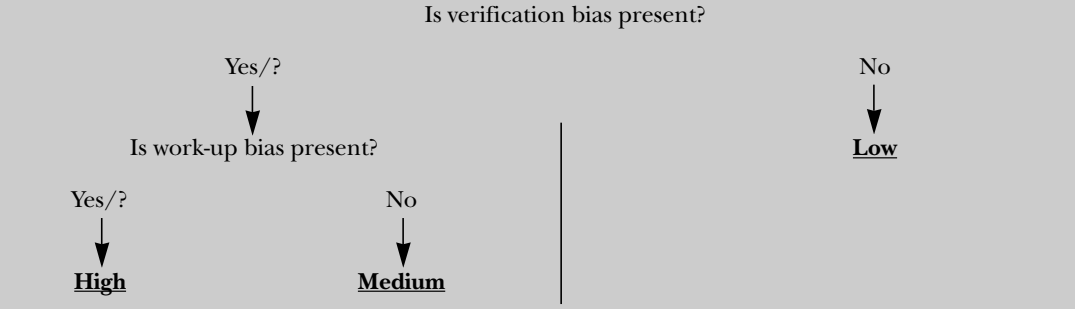
C. Patient cohort bias



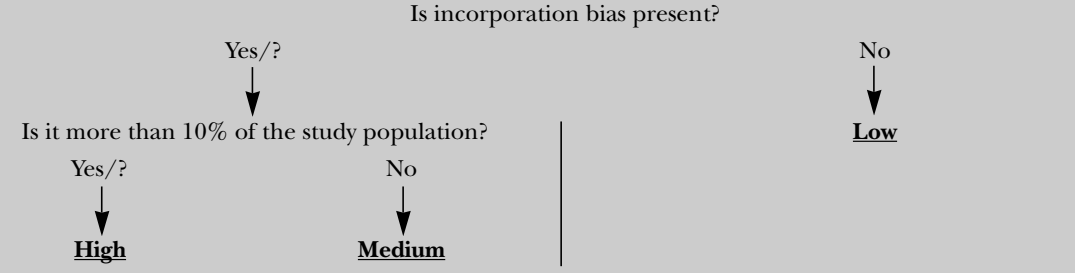
Section 2: Study

Biases associated with application of the gold standard

D. Verification and work-up bias

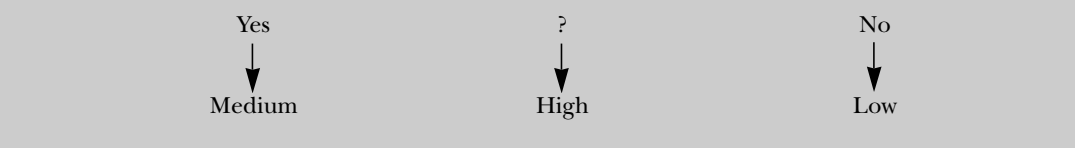


E. Incorporation bias



Biases due to measurement of results

F. Disease progression bias



G. Withdrawal bias

Are results reported for all patients who received the test?

- Yes → **Low**
- No/? → Are there any indeterminate test results?
  - Yes/? → Are there any patients lost to follow-up?
    - Yes/? → **High**
    - No → **Medium**
  - No → **Medium**

H. Observer variability

Is there a single observer of the diagnostic test under evaluation?

- Yes → Is any attempt made to assess intraobserver variability?
  - Yes → **Low**
  - No/? → **Medium**
- No/? → Are results reported separately for each observer?
  - Yes → Is any attempt made to assess intraobserver variability?
    - Yes → **Low**
    - No/? → **Medium**
  - No/? → Is any attempt made to assess interobserver variability?
    - Yes → Is any attempt made to assess intraobserver variability?
      - Yes → **Medium**
      - No/? → **High**
    - No/? → **High**

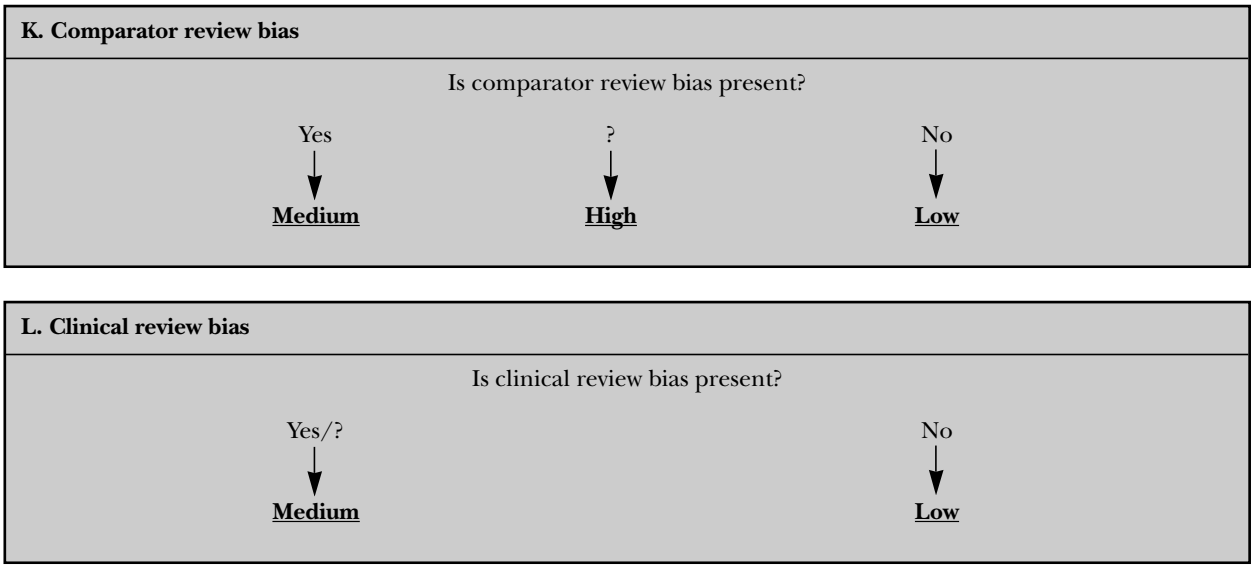
Section 3: Independence of interpretations

Biases associated with blinding between tests

J. Diagnostic and test review bias

Is diagnostic review bias present?

- Yes/? → Is test review bias present?
  - Yes/? → **High**
  - No → **Medium**
- No → Is test review bias present?
  - Yes/? → **Medium**
  - No → **Low**



Step 2: Combining biases into an overall score

The original 20 biases were partly grouped together to form 11 biases, which were assigned a grade of low, medium or high; the high grade corresponds to the likely presence of a bias in a study. Three methods of combining these 11 biases into an overall score are described next, together with the results of applying the system to the 27 EUS studies included in the review of staging performance.

Method 1: Ungrouped, unweighted scoring system

This system did not group the biases any further nor did it weight the biases before combining. Scores were assigned (1 for low, 2 for medium, and 3 for high) for each of the 11 categories of bias, and then simply totalled. The range of scores obtained was 11–33 *(Table A3.3)*. Studies attaining a score of between 11 and 18 inclusive were assigned an overall low score; between 19 and 25 inclusive was scored medium overall; and between 26 and

TABLE A3.3 *Ungrouped, unweighted scoring system*

Grade	Categories	Tally	No. possibilities
High (x 3)	11 High	33	1
	10 High + 1 Medium	32	11
	(10 High + 1 Low) or (9 High + 2 Medium)	31	66
	etc.	30	
		29	
		28	
		27	
		26	
Medium (x 2)		25	
		24	
		23	
	11 Medium	22	
		21	
		20	
		19	
Low (x 1)		18	
		17	
		16	
		15	
	etc.	14	
	(10 Low + 1 High) or (9 Low + 2 Medium)	13	66
	10 Low + 1 Medium	12	11
	11 Low	11	1

33 inclusive was scored high overall. However, as shown in *Table A3.3,* the number of possibilities of obtaining an overall score of low, medium or high is not equal. For example, there is only one way of obtaining the best score of 11 or the worst score of 33, while there are numerous ways of obtaining a medium score of 22. This imbalance is clearly demonstrated in *Figure A3.1,* with the majority of the articles scoring medium.

Method 2: Grouped, unweighted scoring system

The total number of permutations of 11 biases each with a score of high, medium or low is 177,147. By grouping the biases together in subgroups of three, the number of possibilities was reduced to a manageable 27. Hence the 11 biases were grouped into sets of three, as illustrated in *Table A3.4,* and then combined to achieve a score of high, medium or low by using the permutations shown in *Table A3.5*. For example, if a set of three biases was graded high, medium and low (in any order), this corresponded to a tally of six in *Table A3.5* and the combined grade was medium for that set. By continuing this process to combine the three sets of biases together, a single overall score was attained.

For the study category, in order to combine the five biases into a single score for the study set, two combinations were required. First, the three biases due to the measurement of results (disease progression bias, withdrawal bias and observer variability bias) were combined using *Table A3.5* to attain a single score. Secondly, this combined grade was grouped with the two biases associated with the application of the gold standard (verification and work-up bias, and incorporation bias) and were combined again to give the overall grade for the study category. The final score for the article was obtained by combining the three main categories of subjects, study and interpretation.

TABLE A3.4 *Grouping of 11 biases into three categories*

Subjects	Study	Interpretation
Referral bias	Verification and work-up bias	Diagnostic and test review bias
Patient filtering bias	Incorporation bias	Comparator review bias
Patient cohort bias	Disease progression bias Withdrawal bias Observer variability bias	Clinical review bias

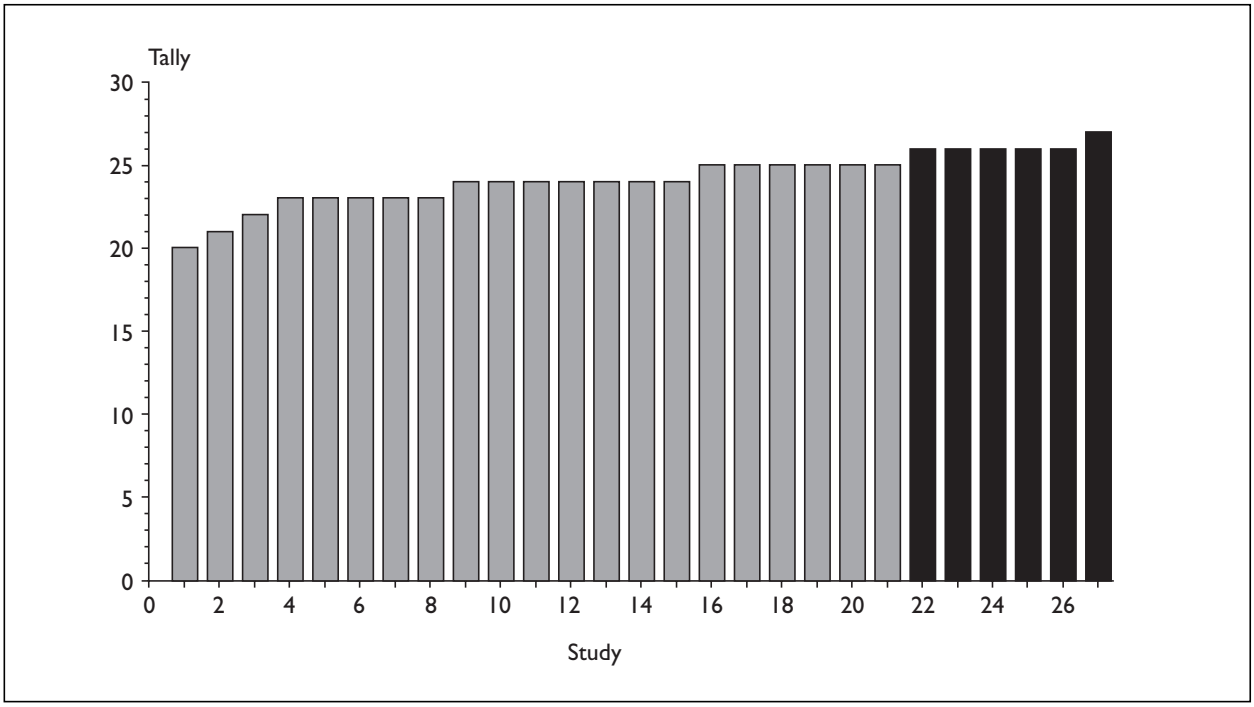


FIGURE A3.1 *Trial of ungrouped, unweighted scoring system (□, medium; ■, high)*

TABLE A3.5 *Grouped, unweighted scoring system*

Grade	a	b	c	Tally	No. possibilities
High (x 3)	H	H	H	9	1
	H	H	M	8	3
	H	M	H	8	
	M	H	H	8	
	H	M	M	7	6
	M	H	M	7	
	M	M	H	7	
	H	H	L	7	
	H	L	H	7	
	L	H	H	7	
Medium (x 2)	M	M	M	6	7
	H	M	L	6	
	H	L	M	6	
	M	H	L	6	
	M	L	H	6	
	L	H	M	6	
	L	M	H	6	
Low (x 1)	M	M	L	5	6
	M	L	M	5	
	L	M	M	5	
	H	L	L	5	
	L	H	L	5	
	L	L	H	5	
	M	L	L	4	3
	L	M	L	4	
	L	L	M	4	
	L	L	L	3	1

H, high; M, medium; L, low

From *Table A3.5* it can be seen that, by reducing the number of biases to be combined from 11 to three, the possible range of scores was reduced from 11–33 to 3–9. The consequence of this is illustrated in the results of scoring the 27 EUS articles *(Figure A3.2)*. The scores tended to group together in clusters and were not particularly well distributed over the range; in fact, all studies scoring medium had a tally of six. The majority of studies using this system scored high for the presence of bias, suggesting an unreasonable standard, which may overemphasise less important biases.

Method 3: Grouped, weighted scoring system

By using the same system for limiting the combination to three categories as in the grouped, unweighted system, a simple alteration was made by weighting the three biases to be combined. Weighting biases can be very advantageous, but it reduces objectivity. A weighting system is useful because it allows important, influential biases significantly to affect the score, and dampens the effect of those less significant biases while still taking them into consideration. The problem arises when deciding which biases are the most important and which are the least. For the purpose of methodology, an arbitrary hierarchy was proposed, as shown in *Table A3.6*.

Using *Table A3.6,* when combining sets of three biases together as described for Method 2, a weight of 3 for high significance, 2 for medium and 1

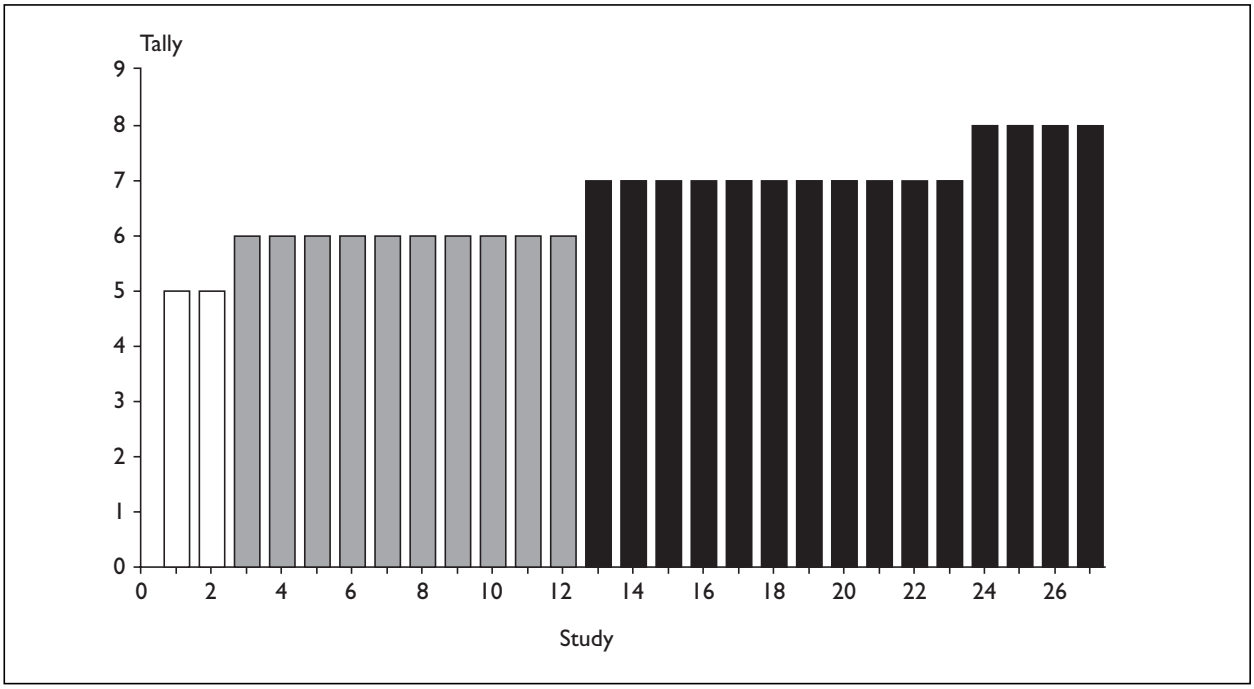


FIGURE A3.2 *Trial of grouped, unweighted scoring system (□, low; ▒, medium; ■, high)*

TABLE A3.6 *Weighting hierarchy of biases*

Section	Bias	Weight
Study	Verification and work-up bias Disease progression bias Withdrawal bias Observer variability Incorporation bias	Most significant ↕ Least significant
Interpretation	Diagnostic and test review bias Comparator review bias Clinical review bias	
Subjects	Patient cohort bias Patient filtering bias Referral bias	

for low significance for each bias was assigned. In this way a bias scoring high and being of high significance will attain a score of 9, while the same bias scoring low will score only 3. For example, the first combination of three for the study section from *Table A3.4* includes disease progression bias, withdrawal bias and observer variability. From these, disease progression bias will be assigned a weight of 3, withdrawal bias a weight of 2, and observer variability a weight of 1. The score from this combination, by using *Table A3.7,* will then be weighted as 2 and combined with verification and work-up bias with a weight of 3, and with incorporation bias with a weight of 1, again using *Table A3.7.* Hence an overall score for the study section will be attained and this will be weighted 3 for the final combination of the overall scores of the three sections.

Table A3.7 illustrates all the possibilities with this weighting system. It is clear that, by weighting the score, the range of possible values has increased, producing a more even spread between low, medium and high (i.e. nine possible ways of scoring low, medium or high). The influence of this balancing of the scores is illustrated in *Figure A3.3,* showing the results of the 27 EUS articles, where a broader range of scores is seen.

Conclusion

The order of importance of biases reported in the weighted methodology was set by the main reviewer. In spite of a number of discussions amongst members of the review panel, it proved impossible to reach a consensus decision on the weighting of the biases and this emphasises the subjective nature of the system. However, the methodology would be valid and reproducible, providing the hierarchy of biases shown in

TABLE A3.7 *Grouped, weighted scoring system*

Grade	a (x 3)	b (x 2)	c (x 1)	Tally	No. possibilities
High (x 3)	H	H	H	18	1
	H	H	M	17	1
	H	M	H	16	2
	H	H	L	16	
	M	H	H	15	2
	H	M	M	15	
	M	H	M	14	3
	H	M	L	14	
	H	L	H	14	
Medium (x 2)	M	H	L	13	3
	M	M	H	13	
	H	L	M	13	
	M	M	M	12	3
	L	H	H	12	
	H	L	L	12	
	L	H	M	11	3
	M	L	H	11	
	M	M	L	11	
Low (x 1)	L	M	H	10	3
	M	L	M	10	
	L	H	L	10	
	L	M	M	9	2
	M	L	L	9	
	L	L	H	8	2
	L	M	L	8	
	L	L	M	7	1
	L	L	L	6	1

H, high; M, medium; L, low

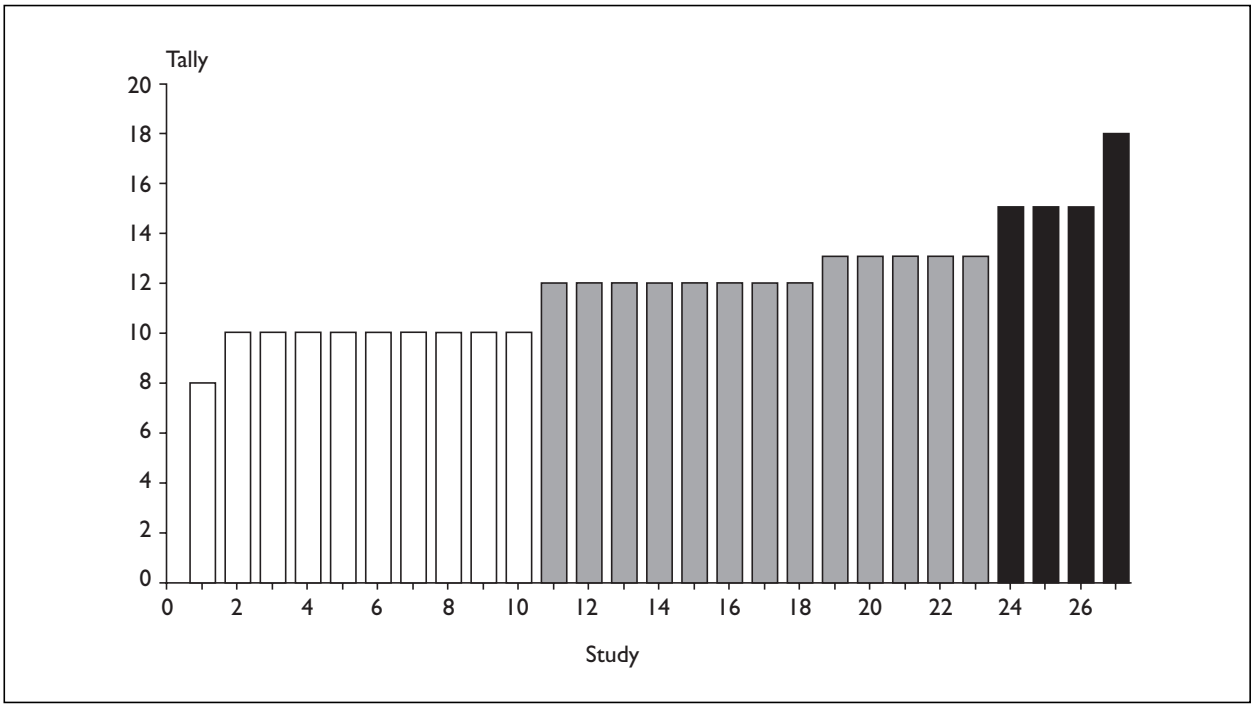


FIGURE A3.3 *Trial of grouped, weighted scoring system (□, low; ▩, medium; ■, high)*

Table A3.6 was acceptable to users. The grouped, weighted system is the best scoring system for producing a broad yet balanced distribution of scores, but the cost is increased subjectivity. Further work is required to generate an objective scoring system using this methodology.

Appendix 4

Checklist results and raw data from primary studies

Contents

A4.1 Checklists

A4.1.1 Oesophageal cancer studies

TABLE A4.1 *Results from bias checklist for oesophageal cancer studies*

	Binmoeller *et al.* 1995[3]	Botet *et al.* 1991[4]	Catalano *et al.* 1994[7]	Dittler *et al.* 1993[9]	Grimm *et al.* 1993[14]	Heintz *et al.* 1991[17]	Hünerbein *et al.* 1996[19]	Manzoni 1993[20]	Murata *et al.* 1988[22]	Murata *et al.* 1993[23]	Peters *et al.* 1994[26]	Takemoto *et al.* 1986[29]	Ziegler *et al.* 1991[35]	Ideal
A1	y	y	y	y	n	n	n	y	y	y	y	y	y	y
A2	n	n	n	n	n	n	n	n	n	n	n	n	n	y
A3	n	n	n	n	n	n	n	n	n	n	n	n	n	y
B1	y	n	y	y	y	n	y	n	n	y	n	y	y	y
B2	n	n	n	n	n	n	n	?	n	n	y	n	n	n
B3	n	n	n	n	n	n	n	n	n	n	n	y	n	n
B4	y	?	?	y	y	?	?	y	?	y	?	n	?	n
C1	n	n	y	y	n	n	n	n	n	n	n	n	n	y
C2	n	y	n	y	n	n	n	n	n	n	n	n	n	y
C3	n	n	n	n	n	n	n	n	n	n	n	n	n	y
D1	y	n	y	y	y	?	y	n	y	n	n	y	y	n
D2	n	n	n	n	n	n	n	n	n	n	n	n	n	n
D3	n	n	n	n	n	n	n	n	n	n	n	n	n	n
E1	?	?	n	?	?	?	?	?	?	?	n	?	n	n
F1	y	y	y	y	y	y	y	n	n	y	n	y	y	y
F2	n	n	n	n	n	n	n	y	y	n	y	n	n	n
F3	n	n	n	n	n	n	n	n	n	n	n	n	n	n
G1	?	?	?	?	?	?	?	y	?	?	?	?	y	y or n
G2	?	?	?	?	?	?	?	y	?	?	?	?	y	y
G3	n	n	n	n	n	n	n	n	n	n	n	n	n	y
G4	?	?	?	?	?	?	?	y	?	?	?	?	y	y or n
G5	n	n	n	n	n	n	n	n	n	n	n	n	n	y
H1	?	?	?	?	?	?	?	?	?	?	?	?	?	n
H2	?	?	?	?	?	?	?	?	?	?	?	?	?	n
H3	n	n	n	n	n	?	n	n	n	n	?	n	?	n
H4	?	?	?	?	?	?	?	?	?	?	?	?	?	n

TABLE A4.2 *Results from factor checklist for oesophageal cancer studies*

		Binmoeller *et al.* 1995[3]	Botet *et al.* 1991[4]	Catalano *et al.* 1994[7]	Dittler *et al.* 1993[9]	Grimm *et al.* 1993[14]	Heintz *et al.* 1991[17]	Hünerbein *et al.* 1996[19]	Manzoni 1993[20]	Murata *et al.* 1988[22]	Murata *et al.* 1993[23]	Peters *et al.* 1994[26]	Takemoto *et al.* 1986[29]	Ziegler *et al.* 1991[35]
Randomised		n	n	n	n	n	n	n	n	n	n	n	n	n
Prospective		y	y	y	y	y	y	y	y	n	y	y	y	y
Controlled		n	n	n	n	n	n	n	y	n	n	n	n	n
No.	T1	0	3	18	27	10	4	2	N/S	57	113	5	8	5
	T2	10	4	16	31	15	5	3	N/S	31	53	6	1	4
	T3	20	20	57	93	30	11	9	7	78	139	23	3	8
	T4	8	23	9	16	8	2	5	8	7	12	0	0	20
Male		(64)	36	72	148	(189)		(146)				(36)	(14)	(37)
Female		(23)	14	28	19	(54)		(105)				(6)	(2)	(15)
Age – years		(61)	60.3	69	56	(59)		(62)				(51)	(63.4)	(57.5)
Age range – years		(40–86)		48–78	29–82			(25–78)				(42–82)	(44–77)	(40–76)
Site	Upper		1		49		5							(4)
	Middle		8		59		8							(30)
	Lower		41		59		9							(18)
Type	Squamous cell carcinoma				108							6		37
	Adenocarcinoma		38		59							36		
	Carcinoma	38	12	100		63		19	24				12	
Model	EUM1									✔			✔	
	EUM2		✔			✔				✔			✔	
	EUM3		✔	✔	✔	✔	✔				✔	✔		
	EUM20			✔										
	JFUM3					✔								
	GFVM2				✔									
	UM-1W										✔			
	E-Probe										✔			
	Pentax FG32U							✔						
	Toshiba linear								✔				✔	
	Siemens linear													✔
	Oesophagoprobe	✔												
	Gold standard	Path	Path	Path	Path/Surg	Path/Surg	Path/Surg	Path	Path	Path	Path	Path	Path	Path/Surg

Brackets indicate data supplied for the whole study group; data were not available for the smaller group that received gold standard verification

A4.1.2 Gastric cancer studies

TABLE A4.3 *Results from bias checklist for gastric cancer studies*

	Akahoshi *et al.* 1991[1]	Botet *et al.* 1991[5]	Caletti *et al.* 1993[6]	Dittler *et al.* 1993[10]	Grimm *et al.* 1993[14]	Hünerbein *et al.* 1996[19]	Massari *et al.* 1996[21]	Murata *et al.* 1988[22]	Perng *et al.* 1996[25]	Saito *et al.* 1991[27]	Shimizu *et al.* 1994[28]	Tio *et al.* 1989[30]	Ziegler *et al.* 1993[36]	Ideal
A1	y	n	y	y	n	n	y	y	y	y	n	y	y	y
A2	n	n	n	n	n	n	n	n	n	n	n	n	n	y
A3	n	n	n	n	n	n	n	n	n	n	n	n	n	y
B1	n	n	n	y	y	y	y	n	y	n	n	n	n	y
B2	?	n	n	n	n	n	n	n	n	?	?	n	n	n
B3	n	n	n	n	n	n	n	n	n	n	n	n	n	n
B4	?	?	?	y	y	?	y	?	n	?	?	?	?	n
C1	y	y	n	y	n	n	y	n	y	n	n	y	y	y
C2	n	y	n	y	n	n	n	n	y	n	n	n	y	y
C3	n	n	y	n	n	n	n	n	y	y	n	n	n	y
D1	n	n	n	y	y	y	y	?	n	n	n	y	n	n
D2	n	n	n	n	n	n	n	n	n	n	n	n	n	n
D3	n	n	n	n	n	n	n	n	n	n	n	n	n	n
E1	?	?	?	?	?	?	?	?	n	?	?	n	n	n
F1	y	y	n	y	n	y	y	y	y	y	y	y	y	y
F2	n	n	y	n	?	n	n	n	n	n	n	n	n	n
F3	n	n	n	n	?	n	n	n	n	n	n	n	n	n
G1	?	?	?	?	?	?	?	?	?	?	?	?	?	y or n
G2	?	?	?	?	?	?	?	?	?	?	?	?	?	y
G3	n	n	n	n	n	n	n	n	n	n	n	n	n	y
G4	?	?	?	?	?	?	?	?	?	?	?	?	?	y or n
G5	n	n	n	n	n	n	n	n	n	n	n	n	n	y
H1	?	n	?	?	?	?	?	?	?	?	?	?	?	n
H2	?	n	?	?	?	?	?	?	?	?	?	?	?	n
H3	?	n	n	n	n	n	?	n	?	n	n	n	?	n
H4	?	?	?	?	?	?	?	?	?	?	?	?	?	n

TABLE A4.4 *Results from factor checklist for gastric cancer studies*

		Akahoshi *et al.* 1991[1]	Botet *et al.* 1991[5]	Caletti *et al.* 1993[6]	Dittler *et al.* 1993[10]	Grimm *et al.* 1993[14]	Hünerbein *et al.* 1996[19]	Massari *et al.* 1996[21]	Murata *et al.* 1988[22]	Perng *et al.* 1996[25]	Saito *et al.* 1991[27]	Shimizu *et al.* 1994[28]	Tio *et al.* 1989[30]	Ziegler *et al.* 1993[36]
Randomised		n	n	n	n	n	n	n	n	n	n	n	n	n
Prospective		y	y	y	y	y	y	y	n	y	y	y	y	y
Controlled		n	n	n	n	n	n	n	n	n	n	n	n	n
No.	T1	40	4	6	27	38	12	12	91	22	46	80	13	22
	T2	21	8	5	52	56	16	14	14	8	14	10	15	32
	T3	8	31	14	151	33	23	21	41	14	45	38	32	36
	T4	5	7	10	24	20	9	18	0	23	5	0	8	18
Male		49	26	(28)	165	(122)	(146)	(53)		(40)			(45)	58
Female		25	24	(14)	89	(81)	(105)	(46)		(29)			(27)	50
Age – years		60	61	(63)	62	(61)	(62)	(61)		(56)			(61.3)	58
Age range – years		17–85	33–81	(35–77)	28–79		(25–78)	(24–79)		(28–72)			(13–87)	29–82
Site	Antrum		12		42									28
	Corpus		21		36									43
	Fundus		13		28									37
	Whole				2									
Type	Adenocarcinoma		46		108					67				101
	Linitis plastica		4											
	Signet ring ca.													7
	Carcinoma					147	60						68	
	Cancer	74		35				65	146		110	128		
Model	EUM1								✔				✔	
	EUM2	✔	✔	✔		✔			✔		✔		✔	
	EUM3	✔	✔	✔	✔	✔		✔		✔	✔	✔	✔	✔
	JFUM3					✔								
	GFVM2				✔									
	EUM20											✔		
	Pentax FG32U						✔							
Gold standard		Path	Path	Path/Surg	Path/Surg	Path/Surg	Path	Path	Path	Path	Path	Path	Path	Path

Brackets indicate data supplied for the whole study group; data were not available for the smaller group that received gold standard verification

A4.1.3 Cardia or gastro-oesophageal junction cancer studies

TABLE A4.5 *Results from bias checklist for cardia or gastro-oesophageal junction cancer studies*

	Altorki *et al.* 1996[2]	François *et al.* 1996 [11]	Greenberg *et al.* 1994[13]	Hordijk *et al.* 1993[18]	Ideal
A1	n	n	y	y	y
A2	n	n	n	n	y
A3	n	n	n	n	y
B1	n	n	y	y	y
B2	?	?	n	n	n
B3	n	n	n	n	n
B4	?	?	n	?	n
C1	n	y	n	n	y
C2	y	y	n	n	y
C3	n	n	n	n	y
D1	n	y	y	y	n
D2	n	n	n	n	n
D3	n	n	n	n	n
E1	?	?	?	n	n
F1	y	y	y	y	y
F2	n	n	n	n	n
F3	n	n	n	n	n
G1	n	?	?	?	y or n
G2	n	?	?	?	y
G3	n	n	n	n	y
G4	?	?	?	?	y or n
G5	n	n	n	n	y
H1	?	?	?	?	n
H2	?	?	?	?	n
H3	n	n	?	?	n
H4	?	?	?	?	n

TABLE A4.6 *Results from factor checklist for cardia or gastro-oesophageal junction cancer studies*

		Altorki *et al.* 1996[2]	François *et al.* 1996 [11]	Greenberg *et al.* 1994[13]	Hordijk *et al.* 1993[18]
Randomised		n	n	n	n
Prospective		y	y	y	y
Controlled		n	n	n	n
No.	T1	11	8	0	9
	T2	8	4	6	3
	T3	34	9	12	28
	T4	2	8	2	1
Male		N/S	24	(23)	(47)
Female		N/S	5	(5)	(15)
Age – years		N/S	65.8	(68)	(62)
Age range – years		N/S	38–84	(51–83)	(35–80)
Site	Oesophagus	✔		✔	✔
	Cardia		29		
	Junction	✔		✔	✔
Type	Adenocarcinoma	36	29		23
	Squamous cell carcinoma	19			18
	Carcinoma			20	
Model	EUM2	✔			
	EUM3	✔	✔	✔	✔
Gold standard		Path	Path	Path/Surg	Path/Surg

Brackets indicate data supplied for the whole study group; data were not available for the smaller group that received gold standard verification

A4.2 Raw data – tumour staging

A4.2.1 Raw data – oesophageal tumour staging

TABLE A4.7 *Results of staging T1 oesophageal cancer*

T1	Binmoeller *et al.* 1995[3]	Botet *et al.* 1991[4]	Catalano *et al.* 1994[7]	Dittler *et al.* 1993[9]	Grimm *et al.* 1993[14]	Heintz *et al.* 1991[17]	Hünerbein *et al.* 1996[19]	Manzoni 1993[20]	Murata *et al.* 1988[22]	Murata *et al.* 1993[23]	Peters *et al.* 1994[26]	Takemoto *et al.* 1986[29]	Ziegler *et al.* 1991[35]
TP	N/S	2	6	22	9	2	0	N/S	52	97	1	6	4
FN	N/S	1	12	5	1	2	2	N/S	4	16	4	2	1
FP	N/S	0	2	3	1	0	0	N/S	3	5	1	0	1
TN	N/S	47	80	137	52	18	17	N/S	113	199	28	4	31
Sensitivity	N/A	66.7	33.3	81.5	90.0	50.0	0.0	N/A	92.9	85.8	20.0	75.0	80.0
Specificity	N/A	100.0	97.6	97.9	98.1	100.0	100.0	N/A	97.4	97.5	96.6	100.0	96.9
PPV	N/A	100.0	75.0	88.0	90.0	100.0	N/A	N/A	94.5	95.1	50.0	100.0	80.0
NPV	N/A	97.9	87.0	96.5	98.1	90.0	89.5	N/A	96.6	92.6	87.5	66.7	96.9
Accuracy	N/A	98.0	86.0	95.2	96.8	90.9	89.5	N/A	95.9	93.4	85.3	83.3	94.6
OR	N/A	N/A	20.0	200.9	468.0	N/A	N/A	N/A	489.7	241.3	7.0	N/A	124.0

TABLE A4.8 *Results of staging T2 oesophageal cancer*

T2	Binmoeller *et al.* 1995[3]	Botet *et al.* 1991[4]	Catalano *et al.* 1994[7]	Dittler *et al.* 1993[9]	Grimm *et al.* 1993[14]	Heintz *et al.* 1991[17]	Hünerbein *et al.* 1996[19]	Manzoni 1993[20]	Murata *et al.* 1988[22]	Murata *et al.* 1993[23]	Peters *et al.* 1994[26]	Takemoto *et al.* 1986[29]	Ziegler *et al.* 1991[35]
TP	8	2	12	24	12	4	2	N/S	24	36	4	1	3
FN	2	2	4	7	3	1	1	N/S	7	17	2	0	1
FP	1	1	17	12	2	4	2	N/S	5	22	8	3	1
TN	27	45	67	124	46	13	14	N/S	136	242	20	8	32
Sensitivity	80.0	50.0	75.0	77.4	80.0	80.0	66.7	N/A	77.4	67.9	66.7	100.0	75.0
Specificity	96.4	97.8	79.8	91.2	95.8	76.5	87.5	N/A	96.5	91.7	71.4	72.7	97.0
PPV	88.9	66.7	41.4	66.7	85.7	50.0	50.0	N/A	82.8	62.1	33.3	25.0	75.0
NPV	93.1	95.7	94.4	94.7	93.9	92.9	93.3	N/A	95.1	93.4	90.9	100.0	97.0
Accuracy	92.1	94.0	79.0	88.6	92.1	77.3	84.2	N/A	93.0	87.7	70.6	75.0	94.6
OR	108.0	45.0	11.8	35.4	92.0	13.0	14.0	N/A	93.3	23.3	5.0	N/A	96.0

TABLE A4.9 *Results of staging T3 oesophageal cancer*

T3	Binmoeller *et al.* 1995[3]	Botet *et al.* 1991[4]	Catalano *et al.* 1994[7]	Dittler *et al.* 1993[9]	Grimm *et al.* 1993[14]	Heintz *et al.* 1991[17]	Hünerbein *et al.* 1996[19]	Manzoni 1993[20]	Murata *et al.* 1988[22]	Murata *et al.* 1993[23]	Peters *et al.* 1994[26]	Takemoto *et al.* 1986[29]	Ziegler *et al.* 1991[35]
TP	19	19	47	83	28	9	9	6	74	126	18	2	7
FN	1	1	10	10	2	2	0	1	4	13	5	1	1
FP	3	1	4	4	5	1	1	3	5	14	2	0	2
TN	15	29	39	68	28	10	9	14	89	164	9	9	27
Sensitivity	95.0	95.0	82.5	89.2	93.3	81.8	100.0	85.7	94.9	90.6	78.3	66.7	87.5
Specificity	83.3	96.7	90.7	94.4	84.8	90.9	90.0	82.4	94.7	92.1	81.8	100.0	93.1
PPV	86.4	95.0	92.2	95.4	84.8	90.0	90.0	66.7	93.7	90.0	90.0	100.0	77.8
NPV	93.8	96.7	79.6	87.2	93.3	83.3	100.0	93.3	95.7	92.7	64.3	90.0	96.4
Accuracy	89.5	96.0	86.0	91.5	88.9	86.4	94.7	83.3	94.8	91.5	79.4	91.7	91.9
OR	95.0	551.0	45.8	141.1	78.4	45.0	N/A	28.0	329.3	113.5	16.2	N/A	94.5

TABLE A4.10 *Results of staging T4 oesophageal cancer*

T4	Binmoeller *et al.* 1995[3]	Botet *et al.* 1991[4]	Catalano *et al.* 1994[7]	Dittler *et al.* 1993[9]	Grimm *et al.* 1993[14]	Heintz *et al.* 1991[17]	Hünerbein *et al.* 1996[19]	Manzoni 1993[20]	Murata *et al.* 1988[22]	Murata *et al.* 1993[23]	Peters *et al.* 1994[26]	Takemoto *et al.* 1986[29]	Ziegler *et al.* 1991[35]
TP	7	23	8	14	5	2	5	7	7	12	N/S	N/S	19
FN	1	0	1	2	3	0	0	1	0	0	N/S	N/S	1
FP	0	2	4	3	1	0	0	1	2	5	N/S	N/S	0
TN	30	25	87	148	54	20	14	15	163	300	N/S	N/S	17
Sensitivity	87.5	100.0	88.9	87.5	62.5	100.0	100.0	87.5	100.0	100.0	N/A	N/A	95.0
Specificity	100.0	92.6	95.6	98.0	98.2	100.0	100.0	93.8	98.8	98.4	N/A	N/A	100.0
PPV	100.0	92.0	66.7	82.4	83.3	100.0	100.0	87.5	77.8	70.6	N/A	N/A	100.0
NPV	96.8	100.0	98.9	98.7	94.7	100.0	100.0	93.8	100.0	100.0	N/A	N/A	94.4
Accuracy	97.4	96.0	95.0	97.0	93.7	100.0	100.0	91.7	98.8	98.4	N/A	N/A	97.3
OR	N/A	N/A	174.0	345.3	90.0	N/A	N/A	105.0	N/A	N/A	N/A	N/A	N/A

A4.2.2 Raw data – gastric tumour staging

TABLE A4.11 *Results of staging T1 gastric cancer*

T1	Akahoshi *et al.* 1991[1]	Botet *et al.* 1991[5]	Caletti *et al.* 1993[6]	Dittler *et al.* 1993[10]	Grimm *et al.* 1993[14]	Hünerbein *et al.* 1996[19]	Massari *et al.* 1996[21]	Murata *et al.* 1988[22]	Perng *et al.* 1996[25]	Saito *et al.* 1991[27]	Shimizu *et al.* 1994[28]	Tio *et al.* 1989[30]	Ziegler *et al.* 1993[36]
TP	37	4	5	22	28	4	12	85	14	41	72	10	20
FN	3	0	1	5	10	8	0	6	8	5	8	3	2
FP	7	1	0	1	3	0	2	5	0	2	1	0	2
TN	27	45	29	226	106	48	51	50	45	62	47	55	84
Sensitivity	92.5	100.0	83.3	81.5	73.7	33.3	100.0	93.4	63.6	89.1	90.0	76.9	90.9
Specificity	79.4	97.8	100.0	99.6	97.2	100.0	96.2	90.9	100.0	96.9	97.9	100.0	97.7
PPV	84.1	80.0	100.0	95.7	90.3	100.0	85.7	94.4	100.0	95.3	98.6	100.0	90.9
NPV	90.0	100.0	96.7	97.8	91.4	85.7	100.0	89.3	84.9	92.5	85.5	94.8	97.7
Accuracy	86.5	98.0	97.1	97.6	91.2	86.7	96.9	92.5	88.1	93.6	93.0	95.6	96.3
OR	47.6	N/A	N/A	994.4	98.9	N/A	N/A	141.7	N/A	254.2	423.0	N/A	420.0

TABLE A4.12 *Results of staging T2 gastric cancer*

T2	Akahoshi *et al.* 1991[1]	Botet *et al.* 1991[5]	Caletti *et al.* 1993[6]	Dittler *et al.* 1993[10]	Grimm *et al.* 1993[14]	Hünerbein *et al.* 1996[19]	Massari *et al.* 1996[21]	Murata *et al.* 1988[22]	Perng *et al.* 1996[25]	Saito *et al.* 1991[27]	Shimizu *et al.* 1994[28]	Tio *et al.* 1989[30]	Ziegler *et al.* 1993[36]
TP	12	6	5	37	41	9	12	7	5	9	6	14	26
FN	9	2	0	15	15	7	2	7	3	5	4	1	6
FP	3	1	2	16	11	10	2	6	9	5	7	6	6
TN	50	41	28	186	80	34	49	126	50	91	111	47	70
Sensitivity	57.1	75.0	100.0	71.2	73.2	56.3	85.7	50.0	62.5	64.3	60.0	93.3	81.3
Specificity	94.3	97.6	93.3	92.1	87.9	77.3	96.1	95.5	84.7	94.8	94.1	88.7	92.1
PPV	80.0	85.7	71.4	69.8	78.8	47.4	85.7	53.8	35.7	64.3	46.2	70.0	81.3
NPV	84.7	95.3	100.0	92.5	84.2	82.9	96.1	94.7	94.3	94.8	96.5	97.9	92.1
Accuracy	83.8	94.0	94.3	87.8	82.3	71.7	93.8	91.1	82.1	90.9	91.4	89.7	88.9
OR	22.2	123.0	N/A	28.7	19.9	4.4	147.0	21.0	9.3	32.8	23.8	109.7	50.6

TABLE A4.13 *Results of staging T3 gastric cancer*

T3	Akahoshi *et al.* 1991[1]	Botet *et al.* 1991[5]	Caletti *et al.* 1993[6]	Dittler *et al.* 1993[10]	Grimm *et al.* 1993[14]	Hünerbein *et al.* 1996[19]	Massari *et al.* 1996[21]	Murata *et al.* 1988[22]	Perng *et al.* 1996[25]	Saito *et al.* 1991[27]	Shimizu *et al.* 1994[28]	Tio *et al.* 1989[30]	Ziegler *et al.* 1993[36]
TP	8	30	12	132	28	20	18	38	11	44	36	26	31
FN	0	1	2	19	5	3	3	3	3	1	2	6	5
FP	4	2	1	19	17	11	2	5	7	5	6	2	6
TN	62	17	20	84	97	26	42	100	46	60	84	34	66
Sensitivity	100.0	96.8	85.7	87.4	84.8	87.0	85.7	92.7	78.6	97.8	94.7	81.3	86.1
Specificity	93.9	89.5	95.2	81.6	85.1	70.3	95.5	95.2	86.8	92.3	93.3	94.4	91.7
PPV	66.7	93.8	92.3	87.4	62.2	64.5	90.0	88.4	61.1	89.8	85.7	92.9	83.8
NPV	100.0	94.4	90.9	81.6	95.1	89.7	93.3	97.1	93.9	98.4	97.7	85.0	93.0
Accuracy	94.6	94.0	91.4	85.0	85.0	76.7	92.3	94.5	85.1	94.5	93.8	88.2	89.8
OR	N/A	255.0	120.0	30.7	32.0	15.8	126.0	253.3	24.1	528.0	252.0	73.7	68.2

TABLE A4.14 *Results of staging T4 gastric cancer*

T4	Akahoshi *et al.* 1991[1]	Botet *et al.* 1991[5]	Caletti *et al.* 1993[6]	Dittler *et al.* 1993[10]	Grimm *et al.* 1993[14]	Hünerbein *et al.* 1996[19]	Massari *et al.* 1996[21]	Murata *et al.* 1988[22]	Perng *et al.* 1996[25]	Saito *et al.* 1991[27]	Shimizu *et al.* 1994[28]	Tio *et al.* 1989[30]	Ziegler *et al.* 1993[36]
TP	3	6	10	19	17	6	16	N/S	19	4	N/S	7	16
FN	2	1	0	5	3	3	2	N/S	4	1	N/S	1	2
FP	0	0	0	8	2	0	1	N/S	2	0	N/S	3	1
TN	69	43	25	222	125	51	46	N/S	42	105	N/S	57	89
Sensitivity	60.0	85.7	100.0	79.2	85.0	66.7	88.9	N/A	82.6	80.0	N/A	87.5	88.9
Specificity	100.0	100.0	100.0	96.5	98.4	100.0	97.9	N/A	95.5	100.0	N/A	95.0	98.9
PPV	100.0	100.0	100.0	70.4	89.5	100.0	94.1	N/A	90.5	100.0	N/A	70.0	94.1
NPV	97.2	97.7	100.0	97.8	97.7	94.4	95.8	N/A	91.3	99.1	N/A	98.3	97.8
Accuracy	97.3	98.0	100.0	94.9	96.6	95.0	95.4	N/A	91.0	99.1	N/A	94.1	97.2
OR	N/A	N/A	N/A	105.5	354.2	N/A	368.0	N/A	99.8	N/A	N/A	133.0	712.0

A4.2.3 Raw data – cardia tumour staging

TABLE A4.15 *Results of staging T1 cardia or gastro-oesophageal junction cancer*

T1	Altorki *et al.* 1996[2]	François *et al.* 1996[11]	Greenberg *et al.* 1994[13]	Hordijk *et al.* 1993[18]
TP	7	7	N/S	5
FN	3	1	N/S	4
FP	0	0	N/S	0
TN	43	21	N/S	30
Sensitivity	70.0	87.5	N/A	55.6
Specificity	100.0	100.0	N/A	100.0
PPV	100.0	100.0	N/A	100.0
NPV	93.5	95.5	N/A	88.2
Accuracy	94.3	96.6	N/A	89.7
OR	N/A	N/A	N/A	N/A

TABLE A4.16 *Results of staging T2 cardia or gastro-oesophageal junction cancer*

T2	Altorki *et al.* 1996[2]	François *et al.* 1996[11]	Greenberg *et al.* 1994[13]	Hordijk *et al.* 1993[18]
TP	5	3	4	0
FN	2	1	2	3
FP	11	3	1	0
TN	35	22	13	36
Sensitivity	71.4	75.0	66.7	0.0
Specificity	76.1	88.0	92.9	100.0
PPV	31.3	50.0	80.0	N/A
NPV	94.6	95.7	86.7	92.3
Accuracy	75.5	86.2	85.0	92.3
OR	8.0	22.0	26.0	N/A

TABLE A4.17 *Results of staging T3 cardia or gastro-oesophageal junction cancer*

T3	Altorki *et al.* 1996[2]	François *et al.* 1996[11]	Greenberg *et al.* 1994[13]	Hordijk *et al.* 1993[18]
TP	24	7	11	25
FN	10	2	1	1
FP	2	2	1	7
TN	17	18	7	6
Sensitivity	70.6	77.8	91.7	96.2
Specificity	89.5	90.0	87.5	46.2
PPV	92.3	77.8	91.7	78.1
NPV	63.0	90.0	87.5	85.7
Accuracy	77.4	86.2	90.0	79.5
OR	20.4	31.5	77.0	21.4

TABLE A4.18 *Results of staging T4 cardia or gastro-oesophageal junction cancer*

T4	Altorki *et al.* 1996[2]	François *et al.* 1996[11]	Greenberg *et al.* 1994[13]	Hordijk *et al.* 1993[18]
TP	2	6	2	1
FN	0	2	0	0
FP	2	1	0	1
TN	49	20	18	37
Sensitivity	100.0	75.0	100.0	100.0
Specificity	96.1	95.2	100.0	97.4
PPV	50.0	85.7	100.0	50.0
NPV	100.0	90.9	100.0	100.0
Accuracy	96.2	89.7	100.0	97.4
OR	N/A	60.0	N/A	N/A

A4.3 Raw data – lymph node staging

A4.3.1 Raw data – lymph node staging of primary oesophageal tumours

TABLE A4.19

Oesophagus	Binmoeller *et al.* 1995[3]	Botet *et al.* 1991[4]	Catalano *et al.* 1994[7]	Dittler *et al.* 1993[9]	Grimm *et al.* 1993[14]	Heintz *et al.* 1991[17]	Hünerbein *et al.* 1996[19]	Peters *et al.* 1994[26]	Ziegler *et al.* 1991[35]
TP	26	35	57	85	37	13	13	22	16
FN	3	1	7	29	3	2	1	2	9
FP	5	5	2	16	5	0	1	6	3
TN	4	9	34	37	17	4	2	4	9
Sensitivity	89.7	97.2	89.1	74.6	92.5	86.7	92.9	91.7	64.0
Specificity	44.4	64.3	94.4	69.8	77.3	100.0	66.7	40.0	75.0
PPV	83.9	87.5	96.6	84.2	88.1	100.0	92.9	78.6	84.2
NPV	57.1	90.0	82.9	56.1	85.0	66.7	66.7	66.7	50.0
Accuracy	78.9	88.0	91.0	73.1	87.1	89.5	88.2	76.5	67.6
OR	6.9	63.0	138.4	6.8	41.9	N/A	26.0	7.3	5.3

A4.3.2 Raw data – lymph node staging of primary gastric tumours

TABLE A4.20

Stomach	Botet *et al.* 1991[5]	Dittler *et al.* 1993[10]	Grimm *et al.* 1993[14]	Hünerbein *et al.* 1996[19]	Massari *et al.* 1996[21]	Perng *et al.* 1996[25]	Tio *et al.* 1989[30]	Ziegler *et al.* 1993[36]
TP	31	130	60	24	40	25	36	40
FN	8	53	13	10	13	12	6	18
FP	1	5	9	5	5	8	15	6
TN	10	66	49	15	7	24	15	44
Sensitivity	79.5	71.0	82.2	70.6	75.5	67.6	85.7	69.0
Specificity	90.9	93.0	84.5	75.0	58.3	75.0	50.0	88.0
PPV	96.9	96.3	87.0	82.8	88.9	75.8	70.6	87.0
NPV	55.6	55.5	79.0	60.0	35.0	66.7	71.4	71.0
Accuracy	82.0	77.2	83.2	72.2	72.3	71.0	70.8	77.8
OR	38.8	32.4	25.1	7.2	4.3	6.3	6.0	16.3

A4.3.3 Raw data – lymph node staging of primary tumours at the cardia

TABLE A4.21

Cardia	Altorki *et al.* 1996[2]	François *et al.* 1996[11]	Greenberg *et al.* 1994[13]
TP	22	16	6
FN	15	3	4
FP	6	0	0
TN	12	10	6
Sensitivity	59.5	84.2	60.0
Specificity	66.7	100.0	100.0
PPV	78.6	100.0	100.0
NPV	44.4	76.9	60.0
Accuracy	61.8	89.7	75.0
OR	2.9	N/A	N/A

A4.4 Raw data – staging of metastases

TABLE A4.22

Metastases	Binmoeller *et al.* 1995[3]	Botet *et al.* 1991[4]	Botet *et al.* 1991[5]	François *et al.* 1996[11]	Tio *et al.* 1989[30]
TP	3	5	N/S	1	4
FN	1	15	N/S	3	2
FP	2	0	N/S	0	0
TN	29	30	N/S	25	66
Sensitivity	75.0	25.0	N/A	25.0	66.7
Specificity	93.5	100.0	N/A	100.0	100.0
PPV	60.0	100.0	N/A	100.0	100.0
NPV	96.7	66.7	N/A	89.3	97.1
Accuracy	91.4	70.0	N/A	89.7	97.2
OR	43.5	N/A	N/A	N/A	N/A

A4.5 Raw data – grouped TNM staging

TABLE A4.23

Altorki *et al.* 1996[2]	Pathology				
	I	IIA	IIB	III	Total
EUS					
I	8	N/S	N/S	N/S	N/S
IIA	N/S	3	N/S	N/S	N/S
IIB	N/S	N/S	1	N/S	N/S
III	N/S	N/S	N/S	22	N/S
Total[a]	**9**	**8**	**8**	**30**	**55**

[a] *Totals as given in article; breakdown not stated*

TABLE A4.24

Botet *et al.* 1991[4]	Pathology				
	I	II	III	IV	Total
EUS					
I	2	0	0	0	2
II	1	4	1	0	6
III	0	2	16	3	21
IV	0	0	0	13	13
Total	**3**	**6**	**17**	**16**	**42**

TABLE A4.25

Botet *et al.* 1991[5]	Pathology				
	I	II	III	IV	Total
EUS					
I	3	0	1	0	4
II	0	7	4	0	11
III	0	1	10	3	14
IV	0	0	0	4	4
Total	**3**	**8**	**15**	**7**	**33**

TABLE A4.26

François *et al.* 1996[11]	Pathology				
	I	II	III	IV	Total
EUS					
I	9	0	0	0	9
II	0	2	1	3	6
III	0	0	7	3	10
IV	0	0	1	3	4
Total	**9**	**2**	**9**	**9**	**29**

TABLE A4.27

Tio *et al.* 1989[30]	Pathology						
	IA	IB	II	IIIA	IIIB	IV	Total
EUS							
IA	8	N/S	N/S	N/S	N/S	N/S	N/S
IB	N/S	5	N/S	N/S	N/S	N/S	N/S
II	N/S	N/S	4	N/S	N/S	N/S	N/S
IIIA	N/S	N/S	N/S	5	N/S	N/S	N/S
IIIB	N/S	N/S	N/S	N/S	15	N/S	N/S
IV	N/S	N/S	N/S	N/S	N/S	4	N/S
Total[a]	**11**	**12**	**11**	**11**	**19**	**4**	**68**

[a] *Totals as given in article; breakdown not stated*

A4.6 Raw data – staging impact comparative studies (CT)

A4.6.1 CT tumour staging

TABLE A4.28

T1	Botet *et al.* 1991[4]	Botet *et al.* 1991[5]	Greenberg *et al.* 1994[13]	Heintz *et al.* 1991[17]	Manzoni 1993[20]	Perng *et al.* 1996[25]	Ziegler *et al.* 1991[35]	Ziegler *et al.* 1993[36]
TP	N/S	N/S	N/S	N/S	N/S	6	1	4
FN	N/S	N/S	N/S	N/S	N/S	4	2	12
FP	N/S	N/S	N/S	N/S	N/S	7	0	12
TN	N/S	N/S	N/S	N/S	N/S	38	30	72
Sensitivity	N/A	N/A	N/A	N/A	N/A	60.0	33.3	25.0
Specificity	N/A	N/A	N/A	N/A	N/A	84.4	100.0	85.7
PPV	N/A	N/A	N/A	N/A	N/A	46.2	100.0	25.0
NPV	N/A	N/A	N/A	N/A	N/A	90.5	93.8	85.7
Accuracy	N/A	N/A	N/A	N/A	N/A	80.0	93.9	76.0
OR	N/A	N/A	N/A	N/A	N/A	8.1	N/A	2.0

TABLE A4.29

T1/T2	Botet *et al.* 1991[4]	Botet *et al.* 1991[5]	Greenberg *et al.* 1994[13]	Heintz *et al.* 1991[17]	Manzoni 1993[20]	Perng *et al.* 1996[25]	Ziegler *et al.* 1991[35]	Ziegler *et al.* 1993[36]
TP	2	4	5	5	20	N/S	N/S	N/S
FN	3	2	1	4	5	N/S	N/S	N/S
FP	1	8	12	2	4	N/S	N/S	N/S
TN	34	16	2	11	25	N/S	N/S	N/S
Sensitivity	40.0	66.7	83.3	55.6	80.0	N/A	N/A	N/A
Specificity	97.1	66.7	14.3	84.6	86.2	N/A	N/A	N/A
PPV	66.7	33.3	29.4	71.4	83.3	N/A	N/A	N/A
NPV	91.9	88.9	66.7	73.3	83.3	N/A	N/A	N/A
Accuracy	90.0	66.7	35.0	72.7	83.3	N/A	N/A	N/A
OR	22.7	4.0	0.8	6.9	25.0	N/A	N/A	N/A

TABLE A4.30

T2/T3	Botet *et al.* 1991[4]	Botet *et al.* 1991[5]	Greenberg *et al.* 1994[13]	Heintz *et al.* 1991[17]	Manzoni 1993[20]	Perng *et al.* 1996[25]	Ziegler *et al.* 1991[35]	Ziegler *et al.* 1993[36]
TP	N/S	N/S	N/S	N/S	N/S	11	7	34
FN	N/S	N/S	N/S	N/S	N/S	11	3	32
FP	N/S	N/S	N/S	N/S	N/S	15	10	22
TN	N/S	N/S	N/S	N/S	N/S	18	13	12
Sensitivity	N/A	N/A	N/A	N/A	N/A	50.0	70.0	51.5
Specificity	N/A	N/A	N/A	N/A	N/A	54.5	56.5	35.3
PPV	N/A	N/A	N/A	N/A	N/A	42.3	41.2	60.7
NPV	N/A	N/A	N/A	N/A	N/A	62.1	81.3	27.3
Accuracy	N/A	N/A	N/A	N/A	N/A	52.7	60.6	46.0
OR	N/A	N/A	N/A	N/A	N/A	1.2	3.0	0.6

TABLE A4.31

T3	Botet *et al.* 1991[4]	Botet *et al.* 1991[5]	Greenberg *et al.* 1994[13]	Heintz *et al.* 1991[17]	Manzoni 1993[20]	Perng *et al.* 1996[25]	Ziegler *et al.* 1991[35]	Ziegler *et al.* 1993[36]
TP	16	7	0	8	7	N/S	N/S	N/S
FN	2	11	12	3	8	N/S	N/S	N/S
FP	13	5	2	5	10	N/S	N/S	N/S
TN	9	7	6	6	29	N/S	N/S	N/S
Sensitivity	88.9	38.9	0.0	72.7	46.7	N/A	N/A	N/A
Specificity	40.9	58.3	75.0	54.5	74.4	N/A	N/A	N/A
PPV	55.2	58.3	0.0	61.5	41.2	N/A	N/A	N/A
NPV	81.8	38.9	33.3	66.7	78.4	N/A	N/A	N/A
Accuracy	62.5	46.7	30.0	63.6	66.7	N/A	N/A	N/A
OR	5.5	0.9	0.0	3.2	2.5	N/A	N/A	N/A

TABLE A4.32

T4	Botet *et al.* 1991[4]	Botet *et al.* 1991[5]	Greenberg *et al.* 1994[13]	Heintz *et al.* 1991[17]	Manzoni 1993[20]	Perng *et al.* 1996[25]	Ziegler *et al.* 1991[35]	Ziegler *et al.* 1993[36]
TP	7	3	1	1	9	12	11	8
FN	10	3	1	1	5	11	9	10
FP	1	3	0	1	4	4	4	20
TN	22	21	18	19	36	28	9	62
Sensitivity	41.2	50.0	50.0	50.0	64.3	52.2	55.0	44.4
Specificity	95.7	87.5	100.0	95.0	90.0	87.5	69.2	75.6
PPV	87.5	50.0	100.0	50.0	69.2	75.0	73.3	28.6
NPV	68.8	87.5	94.7	95.0	87.8	71.8	50.0	86.1
Accuracy	72.5	80.0	95.0	90.9	83.3	72.7	60.6	70.0
OR	15.4	7.0	N/A	19.0	16.2	7.6	2.8	2.5

A4.6.2 CT lymph node staging

TABLE A4.33

Nodes	Botet *et al.* 1991[4]	Botet *et al.* 1991[5]	Greenberg *et al.* 1994[13]	Heintz *et al.* 1991[17]	Manzoni 1993[20]	Perng *et al.* 1996[25]	Ziegler *et al.* 1991[35]	Ziegler *et al.* 1993[36]
TP	23	19	5	10	N/S	12	10	26
FN	6	9	5	5	N/S	15	15	32
FP	5	2	2	3	N/S	27	4	21
TN	8	3	4	1	N/S	26	8	29
Sensitivity	79.3	67.9	50.0	66.7	N/A	44.4	40.0	44.8
Specificity	61.5	60.0	66.7	25.0	N/A	49.1	66.7	58.0
PPV	82.1	90.5	71.4	76.9	N/A	30.8	71.4	55.3
NPV	57.1	25.0	44.4	16.7	N/A	63.4	34.8	47.5
Accuracy	73.8	66.7	56.3	57.9	N/A	47.5	48.6	50.9
OR	6.1	3.2	2.0	0.7	N/A	0.8	1.3	1.1

A4.6.3 CT metastases staging

TABLE A4.34

Metastases	Botet *et al.* 1991[4]
TP	12
FN	4
FP	0
TN	26
Sensitivity	75.0
Specificity	100.0
PPV	100.0
NPV	86.7
Accuracy	90.5
OR	N/A

A4.6.4 CT grouped TNM staging

TABLE A4.35

Botet *et al.* 1991[4] Oesophagus	Pathology I or II	III	IV	Total
EUS				
I or II	5	6	0	11
III	3	10	4	17
IV	0	0	12	12
Total	**8**	**16**	**16**	**40**

TABLE A4.36

Botet *et al.* 1991[5] Stomach	Pathology I or II	III	IV	Total
EUS				
I or II	6	4	1	11
III	2	7	4	13
IV	1	3	2	6
Total	**9**	**14**	**7**	**30**

Appendix 5

Economics checklist

A. Type of analysis

A1 Analytical perspective
- [ ] Individual patient
- [ ] Specific institution
- [ ] Target group for specific services
- [ ] Ministry of Health budget
- [ ] Government's budget
- [ ] Community or society

A2 Type of analysis
- [ ] Cost description
- [ ] Cost outcome description
- [ ] Cost-comparison analysis
- [ ] Cost-effectiveness analysis
- [ ] Cost–benefit analysis
- [ ] Cost–utility analysis
- [ ] Cost-minimisation analysis

A3 Is there comparison of two or more alternatives?	Yes	No	?
A4 Are costs of the alternatives examined?	Yes	No	?
A5 Are consequences of the alternatives examined?	Yes	No	?

B. Outcome indicator

B1 Type of outcome indicator
- [ ] Intermediate endpoints (e.g. sensitivity)
- [ ] Clinical endpoints (e.g. impact on survival)
- [ ] Patient outcome e.g.
  - [ ] Disease-specific QoL
  - [ ] Generic QoL
  - [ ] Utility
  - [ ] Willingness to pay

B2 Is outcome indicator appropriate for type of analysis?	Yes	No	?

C. Cost analysis

C1 Is there a comprehensive range of costs?	Yes	No	?
C2 Are costs measured as opposed to estimated?	Yes	No	?
C3 Are capital costs considered?	Yes	No	?
C4 Are direct and indirect costs separated?	Yes	No	?
C5 Is discounting used?	Yes	No	NApp
C6 Is there a standardised price base?	Yes	No	?

D. Sensitivity analysis

D1 Is sensitivity analysis carried out?	Yes	No	?
D2 Is it for all variables with an observed distribution of values?	Yes	No	?
D3 Is it for all major assumptions on variables not observed?	Yes	No	?
D4 Is threshold analysis performed?	Yes	No	?

(NApp, not applicable: discounting is not always necessary)

Health Technology Assessment panel membership

This report was identified as a priority by the Diagnostics and Imaging Panel.

Acute Sector Panel

Chair: Professor John Farndon, University of Bristol †

Professor Senga Bond, University of Newcastle-upon-Tyne †

Professor Ian Cameron, Southeast Thames Regional Health Authority

Ms Lynne Clemence, Mid-Kent Health Care Trust †

Professor Francis Creed, University of Manchester †

Professor Cam Donaldson, University of Aberdeen

Mr John Dunning, Papworth Hospital, Cambridge †

Professor Richard Ellis, St James's University Hospital, Leeds

Mr Leonard Fenwick, Freeman Group of Hospitals, Newcastle-upon-Tyne †

Professor David Field, Leicester Royal Infirmary †

Ms Grace Gibbs, West Middlesex University Hospital NHS Trust †

Dr Neville Goodman, Southmead Hospital Services Trust, Bristol †

Professor Mark P Haggard, MRC †

Mr Ian Hammond, Bedford & Shires Health & Care NHS Trust

Professor Adrian Harris, Churchill Hospital, Oxford

Professor Robert Hawkins, University of Bristol †

Dr Gwyneth Lewis, Department of Health †

Dr Chris McCall, General Practitioner, Dorset †

Professor Alan McGregor, St Thomas's Hospital, London

Mrs Wilma MacPherson, St Thomas's & Guy's Hospitals, London

Professor Jon Nicholl, University of Sheffield †

Professor John Norman, University of Southampton

Dr John Pounsford, Frenchay Hospital, Bristol †

Professor Michael Sheppard, Queen Elizabeth Hospital, Birmingham †

Professor Gordon Stirrat, St Michael's Hospital, Bristol

Dr William Tarnow-Mordi, University of Dundee

Professor Kenneth Taylor, Hammersmith Hospital, London

Diagnostics and Imaging Panel

Chair: Professor Mike Smith, University of Leeds †

Professor Michael Maisey, Guy's & St Thomas's Hospitals, London *

Professor Andrew Adam, UMDS, London †

Dr Pat Cooke, RDRD, Trent Regional Health Authority

Ms Julia Davison, St Bartholomew's Hospital, London †

Professor Adrian Dixon, University of Cambridge †

Mr Steve Ebdon-Jackson, Department of Health †

Professor MA Ferguson-Smith, University of Cambridge †

Dr Mansel Hacney, University of Manchester

Professor Sean Hilton, St George's Hospital Medical School, London

Mr John Hutton, MEDTAP International Inc., London

Professor Donald Jeffries, St Bartholomew's Hospital, London †

Dr Andrew Moore, Editor, *Bandolier* †

Professor Chris Price, London Hospital Medical School †

Dr Ian Reynolds, Nottingham Health Authority

Professor Colin Roberts, University of Wales College of Medicine

Miss Annette Sergeant, Chase Farm Hospital, Enfield

Professor John Stuart, University of Birmingham

Dr Ala Szczepura, University of Warwick †

Mr Stephen Thornton, Cambridge & Huntingdon Health Commission

Dr Gillian Vivian, Royal Cornwall Hospitals Trust †

Dr Jo Walsworth-Bell, South Staffordshire Health Authority †

Dr Greg Warner, General Practitioner, Hampshire †

Methodology Panel

Chair: Professor Martin Buxton, Brunel University †

Professor Anthony Culyer, University of York *

Dr Doug Altman, Institute of Health Sciences, Oxford †

Professor Michael Baum, Royal Marsden Hospital

Professor Nick Black, London School of Hygiene & Tropical Medicine †

Professor Ann Bowling, University College London Medical School †

Dr Rory Collins, University of Oxford

Professor George Davey-Smith, University of Bristol

Dr Vikki Entwistle, University of Aberdeen †

Professor Ray Fitzpatrick, University of Oxford †

Professor Stephen Frankel, University of Bristol

Dr Stephen Harrison, University of Leeds

Mr John Henderson, Department of Health †

Mr Philip Hewitson, Leeds FHSA

Professor Richard Lilford, Regional Director, R&D, West Midlands †

Mr Nick Mays, King's Fund, London †

Professor Ian Russell, University of York †

Professor David Sackett, Centre for Evidence Based Medicine, Oxford †

Dr Maurice Slevin, St Bartholomew's Hospital, London

Dr David Spiegelhalter, Institute of Public Health, Cambridge †

Professor Charles Warlow, Western General Hospital, Edinburgh †

* Previous Chair
† Current members

continued

continued

Pharmaceutical Panel

Chair: Professor Tom Walley, University of Liverpool †

Professor Michael Rawlins, University of Newcastle-upon-Tyne*

Dr Colin Bradley, University of Birmingham

Professor Alasdair Breckenridge, RDRD, Northwest Regional Health Authority

Ms Christine Clark, Hope Hospital, Salford †

Mrs Julie Dent, Ealing, Hammersmith & Hounslow Health Authority, London

Mr Barrie Dowdeswell, Royal Victoria Infirmary, Newcastle-upon-Tyne

Dr Tim Elliott, Department of Health †

Dr Desmond Fitzgerald, Mere, Bucklow Hill, Cheshire

Dr Felicity Gabbay, Transcrip Ltd †

Dr Alistair Gray, Health Economics Research Unit, University of Oxford †

Professor Keith Gull, University of Manchester

Dr Keith Jones, Medicines Control Agency

Professor Trevor Jones, ABPI, London †

Ms Sally Knight, Lister Hospital, Stevenage †

Dr Andrew Mortimore, Southampton & SW Hants Health Authority †

Mr Nigel Offen, Essex Rivers Healthcare, Colchester †

Dr John Posnett, University of York

Mrs Marianne Rigge, The College of Health, London †

Mr Simon Robbins, Camden & Islington Health Authority, London †

Dr Frances Rotblat, Medicines Control Agency †

Mrs Katrina Simister, Liverpool Health Authority †

Dr Ross Taylor, University of Aberdeen †

Dr Tim van Zwanenberg, Northern Regional Health Authority

Dr Kent Woods, RDRD, Trent RO, Sheffield †

Population Screening Panel

Chair: Professor Sir John Grimley Evans, Radcliffe Infirmary, Oxford †

Dr Sheila Adam, Department of Health*

Ms Stella Burnside, Altnagelvin Hospitals Trust, Londonderry †

Dr Carol Dezateux, Institute of Child Health, London †

Dr Anne Dixon Brown, NHS Executive, Anglia & Oxford †

Professor Dian Donnai, St Mary's Hospital, Manchester †

Dr Tom Fahey, University of Bristol †

Mrs Gillian Fletcher, National Childbirth Trust †

Professor George Freeman, Charing Cross & Westminster Medical School, London

Dr Mike Gill, Brent & Harrow Health Authority †

Dr JA Muir Gray, RDRD, Anglia & Oxford RO †

Dr Anne Ludbrook, University of Aberdeen †

Professor Alexander Markham, St James's University Hospital, Leeds †

Professor Theresa Marteau, UMDS, London

Dr Ann McPherson, General Practitioner, Oxford †

Professor Catherine Peckham, Institute of Child Health, London

Dr Connie Smith, Parkside NHS Trust, London

Dr Sarah Stewart-Brown, University of Oxford †

Ms Polly Toynbee, Journalist †

Professor Nick Wald, University of London †

Professor Ciaran Woodman, Centre for Cancer Epidemiology, Manchester

Primary and Community Care Panel

Chair: Dr John Tripp, Royal Devon & Exeter Healthcare NHS Trust †

Professor Angela Coulter, King's Fund, London *

Professor Martin Roland, University of Manchester *

Dr Simon Allison, University of Nottingham

Mr Kevin Barton, East London & City Health Authority †

Professor John Bond, University of Newcastle-upon-Tyne †

Ms Judith Brodie, Age Concern, London †

Dr Nicky Cullum, University of York †

Professor Shah Ebrahim, Royal Free Hospital, London

Mr Andrew Farmer, Institute of Health Sciences, Oxford †

Ms Cathy Gritzner, The King's Fund †

Professor Andrew Haines, RDRD, North Thames Regional Health Authority

Dr Nicholas Hicks, Oxfordshire Health Authority †

Professor Richard Hobbs, University of Birmingham †

Professor Allen Hutchinson, University of Sheffield †

Mr Edward Jones, Rochdale FHSA

Professor Roger Jones, UMDS, London

Mr Lionel Joyce, Chief Executive, Newcastle City Health NHS Trust

Professor Martin Knapp, London School of Economics & Political Science

Dr Phillip Leech, Department of Health †

Professor Karen Luker, University of Liverpool

Professor David Mant, NHS Executive South & West †

Dr Fiona Moss, Thames Postgraduate Medical and Dental Education †

Professor Dianne Newham, King's College London

Professor Gillian Parker, University of Leicester †

Dr Robert Peveler, University of Southampton †

Dr Mary Renfrew, University of Oxford

Ms Hilary Scott, Tower Hamlets Healthcare NHS Trust, London †

* Previous Chair
† Current members

Health Technology Assessment 1998; Vol. 2: No. 18

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta

ISSN 1366-5278