Development of the NICE clinical guideline on the diagnosis and management of colorectal cancer

Miss Angeliki Kontoyannis BSc MBChB MRCS

Cardiff University

Submitted for the degree of
Doctor of Medicine (MD)

2014

For my parents

Summary

This thesis examines the evidence base on which the NICE guideline for colorectal cancer was developed.

The information supporting guidelines varies. Six separate studies researching the availability and quality of different types of such information were carried out. Methodology, epidemiology, clinical practice, diagnostic accuracy, therapeutic, and internationally sourced data was examined.

The information was sourced by online data mining, national database queries, systematic reviewing of literature databases, and the observation of the NICE guideline development process through both membership of the guideline development group established to produce the recommendations, and the technical development team supporting its production.

Results show that :

- NICE methodology data is available publicly, is easily accessible online, and has been developed following an internationally accepted guideline quality assessment and development tool.
- Epidemiology data on colorectal cancer is easily accessible and of good quality.
- Data regarding current clinical practice on colorectal cancer collected by national databases has methodological challenges and is not easily accessible.
- Diagnostic accuracy studies are less robustly developed comapared to therapeutic studies. Their design is heterogeneous making the results subject to bias and reporting is inconsistent. Quality assessment when evaluating diagnostic evidence for a guideline ensures clarity with regard to the strength of the recommendations.
- Systematic reviews of therapeutic studies can be inappropriately considered high quality evidence using traditional quality appraisal and evidence classification methods. All outcomes of a study should be considered when assesing study quality and recommendations based on a more holistic grading of the evidence are a more accurate reflection of the evidence.
- Evidence considered for guideline development is international in its nature and the national setting of a study does influence guideline recommendations.

Overall, NICE methodology is of high quality. The research helps identify challenges that the evidence can present as a platform for future improvements.

Table of Contents

Publications and presentations resulting from this research

1. Kontoyannis A, Hargest R. How Guidelines influence modern surgical practice. In: Taylor I, Johnson CD (Editors). Recent Advances in Surgery. New Delhi: Jaypee Brothers Medical Publishers; 2011. p 20-33 (ISBN 978-93-5025-355-7)

2. Kontoyannis A, Hargest R. The importance of understanding guideline methodology and the principles of evidence based medicine. BMJ online rapid response. 2011. http://www.bmj.com/rapid-response/2011/11/03/importance-understanding-guideline-development-methodology-and-principles-

3. Kontoyannis A, O'Connell, S, Berendse S, Poston G, Hargest R. Colorectal liver metastases - which diagnostic imaging modality is best. British Journal of Surgery. 2013;100 (S4):2-49
   Presented at the Joint Conference of the Society for Research and Academic Surgery and the Royal Society of Medicine, RSM London, Jan 2013.

4. National Institute for Health and Clinical Excellence (NICE). Clinical guideline 131 - The diagnosis and management of colorectal cancer. London: NICE;2011. (member of the guideline development group)
   http://www.nice.org.uk/nicemedia/live/13597/56998/56998.pdf
   http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0051713/

Glossary

Adenoma	A benign epithelial neoplasm in which the tumor cells form glands or glandlike structures. Usually well circumscribed, tending to compress rather than infiltrate or invade adjacent tissue.
Allele	Any one of a series of two or more different genes that may occupy the same locus on a specific chromosome Autosomal chromosomes are paired, each autosomal gene is represented twice in normal somatic cells. If the same allele occupies both units of the locus, the individual or cell is homozygous for this allele. If the alleles are different, the individual or cell is heterozygous for both alleles.
Appendix	A wormlike intestinal diverticulum extending from the blind end of the cecum; it varies in length and ends in a blind extremity.
Benign	Denoting the non-malignant character of a neoplasm.
Cancer	General term frequently used to indicate any of various types of malignant neoplasms, most of which invade surrounding tissues, may metastasize to several sites, and are likely to recur after attempted removal and to kill the patient unless adequately treated; especially, any such carcinoma or sarcoma, but, in ordinary usage, especially the former.
Carcinoma	Any of various types of malignant neoplasm derived from epithelial cells, chiefly glandular (adenocarcinoma) or squamous (squamous cell carcinoma); the most commonly occurring kind of cancer.
Colon	The large intestine extending from the caecum to the rectum.
DNA	Deoxyribonucleic acid (DNA) is the type of nucleic acid containing deoxyribose as the sugar component and found principally in the nuclei (chromatin, chromosomes) and mitochondria of animal and plant cells. Many forms are known, the most commonly described of which is double stranded / double helix. Chromosomes are composed of double-stranded DNA.
Dysplasia	Abnormal tissue development.

Epigenetic	Relating to epigenesis; the regulation of the expression of gene activity without alteration of genetic structure.
Exophytic	Denoting a neoplasm or lesion that grows outward from an epithelial surface.
Gene	A functional unit of heredity that occupies a specific place (locus) on a chromosome, is capable of reproducing itself exactly at each cell division, and directs the formation of an enzyme or other protein. In organisms reproducing sexually, genes normally occur in pairs in all cells except gametes, as a consequence of the fact that all chromosomes are paired except the sex chromosomes (X and Y) of the male.
Germline	A collection of haploid cells derived from the specialized cells of the primitive gonad.
Growth factor	Natural substances produced by the body (hormones) or obtained from food (vitamins, minerals) that promote growth and development by directing cell maturation and differentiation and by mediating maintenance and repair of tissues; abnormalities in growth factors may be involved in benign and malignant neoplasia.
Malignant	In reference to a neoplasm, having the property of locally invasive and destructive growth and metastasis.
Mass	A lump or aggregation of coherent material. Commonly used as a synonym for tumor or neoplasm.
Metastatic	Relating to metastasis; the spread of a disease process from one part of the body to another, as in the appearance of neoplasms in parts of the body remote from the site of the primary tumour; results from dissemination of tumour cells by the lymphatics or blood vessels or by direct extension through serous cavities or subarachnoid or other spaces.
Mismatch repair	Replacement of mismatched base pairs by removal of the incorrect base and replacement with the correct base by DNA polymerase.
Mucosa	A mucous tissue lining various tubular structures consisting of epithelium, lamina propria, and, in the digestive tract, a layer of smooth muscle (muscularis mucosa).

Mutation	A change in the chemistry of a gene that is perpetuated in subsequent divisions of the cell in which it occurs; a change in the sequence of base pairs in the chromosome.
Neoplasia / Neoplasm	An abnormal tissue that grows by cellular proliferation more rapidly than normal and continues to grow after the stimuli that initiated the new growth cease. Neoplasms show partial or complete lack of structural organization and functional coordination with the normal tissue, and usually form a distinct mass of tissue that may be either benign (benign tumour) or malignant (cancer).
Oncogenes	Any of a family of genes that normally encodes proteins that are involved in cell growth or regulation but that may foster malignant processes if mutated.
Pedunculated polyp	Any form of polyp that is attached to the base tissue by means of a slender stalk.
Polyp	A general descriptive term used with reference to any mass of tissue that bulges or projects outward or upward from the normal surface level, thereby being macroscopically visible as a hemispheroidal, spheroidal, or irregular moundlike structure growing from a relatively broad base or a slender stalk; polyps may be neoplasms, foci of inflammation, degenerative lesions, or malformations.
Rectum	The terminal portion of the digestive tube, extending from the rectosigmoid junction to the anal canal.
Sessile polyp	Any form of polyp that has a relatively broad base.
Somatic cells	The cells of an organism, other than the germ cells.
Stricture	A circumscribed narrowing of a hollow structure.
Tumour suppressor gene	A gene that encodes a protein involved in controlling cellular growth; inactivation of this type of gene leads to deregulated cellular proliferation, as in cancer.

Glossary terms defined using Stedman's Medical Dictionary. London: Lippincott Williams & Wilkins; 2006 via www.medilexicon.com
Accessed 5 Feb 2014

Abbreviations

ACPGBI	Association of Coloproctology of Great Britain and Ireland
AGREE	Appraisal of Guidelines for Research and Evaluation Instrument - original version and version II (AGREE II)
AHRQ	Agency for Healthcare Research and Quality (of the United States)
ALP	Alkaline phosphatase
ALT	Alanine transaminase
APC	Adenomatous polyposis coli
APR	Abdominoperineal resection
ASA	American Society of Anesthesiologists
ÄZQ	German Agency for Quality in Medicine (Ärztliches Zentrum für Qualität in der Medizin)
BAX	B cell lymphoma (BCL)- 2-associated X protein
BC	Before Christ
BRAF	B homolog rapidly accelerated fibrosarcoma proto-oncogene
CANISC	Cancer Network Information System Cymru
CBO	Dutch Institute for Healthcare Improvement (founded as the Central Accompagnement Organization for Peer Review by the Dutch Association of Chief Medical Officers)
CCP	Centre for Clinical Practice at the National Institute of Health and Clinical Excellence
CDKN2A	Cyclin-dependent kinase inhibitor 2A gene; (methylated tumor suppressor gene *p16*)
c-e	Contrast-enhanced
CEA	Carcinoembryonic antigen (tumour marker)
CECT	Contrast-enhanced x-ray computed tomography
CI	95% confidence interval
CIN	Chromosomal instability
CNS	Cancer nurse specialist
COSD	Cancer Outcomes and Services Dataset
CRC	Colorectal cancer

CRM	Circumferential margin
CRUK	Cancer Research UK
CT	X-ray computed tomography
CTAP	X-ray computed tomography during arterioportography
cTcNcM	Clinical staging using the T-N-M classification
CUP	Cancer of unknown primary origin
DCC	Deleted in colon cancer
DCO	Death certificate only: cases represented in colorectal cancer data and identified only from death certificates
DFS	Disease-free survival
DIPEX	Database of Individual Patient Experience
DNA	Deoxyribonucleic acid
DoH	Department of Health
DWI	Diffusion-weighted imaging – a magnetic resonance imaging modality in which the intensity of each three-dimensional image element (voxel) reflects the rate of water diffusion at the corresponding location
DWI MRI	Diffusion-weighted imaging modality of magnetic resonance imaging
EBM	Evidence-based medicine
EGFR	Epidermal growth factor receptor
ENCORE	English National Cancer Online Registration Environment
ESD	Endoscopic submucosal dissection
EUS	Endo-anal ultrasound scan
FAP	Familial adenomatous polyposis
FDG (F-18-FDG)	^{18}F-fluoro-2-deoxy-D-glucose – a radiopharmaceutical and biologically active tracer used in the medical imaging modality positron emission tomography
FDG PET	Positron emission tomography using the tracer ^{18}F-fluorodeoxy-D-glucose (FDG)
FDG PETCT	Positron emission tomography and x-ray computed tomography using the tracer ^{18}F-fluorodeoxy-D-glucose (FDG)
Gad MRI	Gadolinium contrast-enhanced magnetic resonance imaging

GAIN	Guidelines Audit and Implementation Network
GDG	National Institute of Health and Clinical Excellence guideline development group of health professionals
GI	gastrointestinal
GILDA	Gruppo Italiano di Lavoro per la Diagnosi Anticipata – a large-scale multi-center European study to delineate the optimal surveillance strategy following resection of primary colorectal cancer
G-I-N	Guidelines International Network
GMC	General Medical Council
GP(s)	General Practitioner(s)
GPRD	General Practice Research Database
GRADE	Grading of Recommendations Assessment Development and Evaluation – a method of quality assessment and evidence grading developed by the GRADE Working Group commencing in 2000
HES	Hospital Episode Statistics - a data repository containing details of all admissions, out-patient appointments and Accident & Emergency attendances at NHS hospitals in England
hMLH1	Human mutational homolog 1
hMSH2	Human mutational homolog 2
HNPCC	Hereditary non-polyposis colon cancer
HSW	Health Solutions Wales (predecessor of the NHS Wales Informatics Service (NWIS))
HTA	Health Technology Assessment programme of the National Institute for Health Research
IACR	International Association of Cancer Registries
IARC	International Agency for Research on Cancer (of the World Health Organization)
ICBP	International Cancer Benchmarking Partnership
ICD-10	International Classification of Diseases version 10
ICER	Incremental cost-effectiveness ratio

IOUS	Intra-operative ultrasound
IP	Interventional procedures
IT	Index test(s)
I^2 (I squared)	I squared index is the percentage of total variation across studies in a meta-analysis that is due to heterogeneity rather than chance
K-ras	*Kirsten* rat sarcoma viral oncogene homolog
lapUSS	Laparoscopic ultrasound scan
LETR	'Linking Evidence To Recommendation'
LSHTM	London School of Hygiene and Tropical Medicine
MDCT	Multidetector row x-ray computed tomography
MDT	Multi-disciplinary team
MeSH	Medical Subject Headings - a thesaurus of medical terms used by many databases and libraries to index and classify medical information. It helps to overcome differences in UK and US English and different terminology applied to identical concepts
met	Solitary metastasis
mets	Metastases
MLH1	Mutational homolog 1
MMR	Mismatch-repair
MnDPDP	Mangafodipir trisodium contrast-enhanced magnetic
MRI	resonance imaging
MOSAIC	Multicentre international study of oxaliplatin, 5-fluorouracil and leucovorin in the adjuvant treatment of colon cancer
MRI	Magnetic resonance imaging
MSI	Microsatellite instability
NA	Not applicable
NATCANSAT	National Cancer Services Analysis Team
NBCA	National Bowel Cancer Audit
NBOCAP	National Bowel Cancer Audit Project
NCASP	National Clinical Audit Support Programme of the NHS Information Centre (NHS-IC)

NCC(s)	National Collaborating Centre(s)
NCC-C	National Collaborating Centre for Cancer
NCCN	National Comprehensive Cancer Network
NCDR	National Cancer Data Repository
NCIN	National Cancer Intelligence Network
NCRI	National Cancer Research Institute
n-e	Non-enhanced
NGC	National Guideline Clearinghouse (of the United States)
NHG	Dutch College of General Practitioners (Nederlands Huisartsen Genootschap)
NHS	National Health Service
NHS-IC	NHS Information Centre
NWIS	NHS Wales Informatics Service
N-H-L	Non-Hodgkin lymphoma
NICE	National Institute of Health and Clinical Excellence
NICR	Northern Ireland Cancer Registry
NMSC	Non-melanoma skin cancer
No. / no.	number of
NPV	Negative predictive value
NSABP C-07	National Surgical Adjuvant Breast and Bowel Project study 07 to evaluate the efficacy of different adjuvant chemotherapy regimens for Dukes stages B and C colon cancer
NYCRIS	Northern and Yorkshire Cancer Registry and Information Service
ONS	Office for National Statistics
OPCS-4.5	Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures version 4.5
OR	Odds ratio
OS	Overall survival
p	P-value: in statistical significance testing, the probability of obtaining a test statistic at least as extreme as the one observed assuming the null hypothesis is true
PbR	‘Payment by results’ policy for the remuneration of health

	service providers
PEDW	Patient Episode Database for Wales
PET	Positron emission tomography
PETCT	Positron emission tomography and x-ray computed tomography
PICO	Patient, intervention, comparison and outcome framework
PPV	Positive predictive value
pre-op	Pre-operative
post-op	Post-operative
pTpNpM	Pathological staging using the T-N-M classification
p14	Tumor suppressor gene *p14*
p16	Tumor suppressor gene *p16*
p53	Tumor protein p53, in humans encoded by the *TP53* gene
QALY(s)	Quality Adjusted Life Year(s)
QoL	Quality of Life
QUADAS	Quality Assessment of Diagnostic Accuracy Studies (quality assessment tool)
RCT(s)	Randomised controlled trial(s)
REVMAN 5	Review Manager 5: the Cochrane Collaboration's meta-analysis software for preparing and maintaining Cochrane reviews (version 5)
RS	Reference standard
RTDS	Radiotherapy dataset
R0	Resection for cure or complete remission in the R classification adopted in 1987 by the Union for International Cancer Control (UICC) denoting the absence or presence of residual tumour after treatment
ROC	Receiver-operator characteristic curve
SACT	Systemic Anti-Cancer Therapy Data Standard
SBRT	Stereotactic body radiotherapy
SCPRT	Short-course pre-operative radiotherapy
SCR	Scottish Cancer Registry
SEMS	Self-expanding metal stent(s)

SIGN	Scottish Intercollegiate Guidelines Network
SPIO-MRI	Superparamagnetic iron oxide nanoparticle contrast-enhanced magnetic resonance imaging
spiral CT	Spiral (or helical) cone beam x-ray computed tomography
SR(s)	Systematic review(s)
SSCRGs	Site-Specific Clinical Reference Groups
SS-EPI-DWI	Single shot echo planar diffusion-weighted magnetic resonance imaging
SS SE-EPI	Unenhanced single-shot spin-echo echo planar magnetic resonance imaging
SUS	'Secondary User Services' (*British Telecom* database management system)
TEMS	Transanal endoscopic microsurgery
TGF-β	Transforming growth factor beta
TGFBR2	Transforming growth factor beta receptor type 2
TNM	Tumor–node–metastases
TN	True negative(s)
TP	True positive(s)
UICC	Union for International Cancer Control
UK	United Kingdom
UKACR	United Kingdom Association of Cancer Registries
US	United States
USA	United States of America
USS	Ultrasound scan
v	Versus
WICSU	Welsh Cancer Intelligence and Survival Unit
XNMSC	Excluding non-melanoma skin cancer
1°	Primary
15-PGDH	15-prostaglandin dehydrogenase

List of figures and tables.

Figure

Acknowledgments

I am most sincerely grateful to Miss Rachel Hargest, Senior Lecturer at Cardiff University and Consultant Colorectal Surgeon at the University Hospital of Wales, Cardiff who was exemplary in her role as my main supervisor. She was patient and supportive throughout this period of research that coincided with a number of challenging life events for me. She worked tirelessly to secure funding and provided guidance and instruction of superior quality whenever needed.

I am also grateful to Dr Andy Champion at the National Collaborating Centre for Cancer in Cardiff. He was supportive and guided me patiently, with a contagious enthusiasm for the discipline of evidence-based medicine.

This project could not have started without the vision of Dr Fergus McBeth Director of guidelines at the National Institute of Clinical Excellence and previously Director of the National Collaborating Centre for Cancer in Cardiff and for this I am very grateful to him.

I am also very grateful to the chair of the guideline development group, Mr Graeme Poston and all the members of the group, for making me feel welcome, encouraging me, and for their advice and support.

I would also like to thank all at the NCC-C for sharing their expertise. Director Dr John Graham, Assistant Manager Miss Angela Bennett, Reviewers Dr Angela Melder, Dr Sue O' Connell, Dr Mia Schmidt-Hansen, Dr Nathan Bromham, Information specialist Dr Sabine Berendse, health economist Dr Bernadette Li.

I would also like to thank Mr Brian Rees, Miss Helen Sweetland, Mr Malcolm Puntis for their continued and valued mentorship, as well as Prof Geraint Williams, Prof Tim Maughan, Mr Jeff Stamatakis, Dr John Green, Dr Sunil Dolwani for their advice on clinical and research matters.

I am grateful to the London deanery and my surgical training programme director Mr Mike Saunders for granting me time out of clinical training for the purpose of this research.

I am indebted to all who have assisted with funding: Mr David Scott-Coombes at University Hospital Wales, The Aghia Philothei Trust, The Greek Orthodox Charity, ELKA shipping, my parents and my husband.

I owe a special thanks to Dr Rosie Coulson for accommodation while I lived and carried out the research in Cardiff.

Last but not least I am grateful to my family, my mother, my brother, my husband Murdoch and my two babies Joanna and Ronald whose love, selfless support beyond the call of duty, understanding, encouragement and endless patience has allowed me to complete this work.

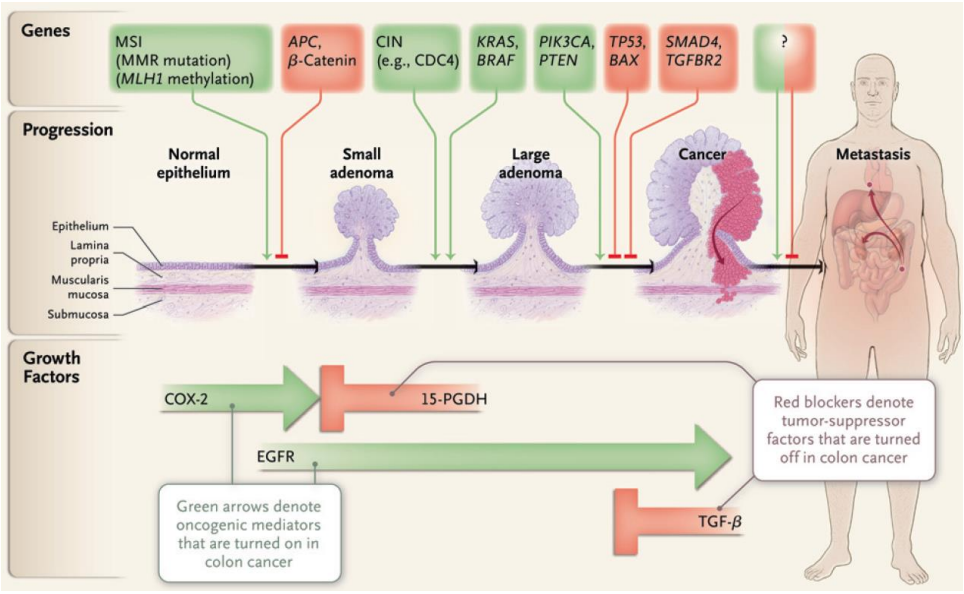
1.0 Introduction

1.1 Colorectal cancer

Colorectal cancer (CRC) is the third most common cancer in the UK and the second leading cause of cancer deaths.[1] Colorectal cancer includes cancerous growths in the colon, rectum and appendix.[1] Most colorectal cancers arise from adenomatous polyps (adenomas). These neoplasms are usually benign, but some develop into cancer over time.[2] It is a common form of malignancy in developed countries but occurs much less frequently in the developing world.[1]

Colon cancer is believed to be caused by a cascade of genetic mutations leading to progressively disordered local DNA replication and accelerated colonocyte replication. The progressive accumulation of multiple genetic mutations results in the transition from normal mucosa to benign adenoma to severe dysplasia to carcinoma. This is referred to as the adenoma-to-cancer sequence and is presented graphically in figure 1.1.[2]

Figure 1.1: Genes & Growth Factor Pathways that drive the progression of colorectal cancer.[3]



In the progression of colon cancer, genetic alterations target the genes that are identified at the top of the figure 1. The microsatellite instability (MSI) pathway is initiated by mismatch-repair (*MMR*) gene mutation or by aberrant *MLH1* methylation and is further associated with downstream mutations in *TGFBR2* and *BAX*. Aberrant *MLH1* methylation and *BRAF* mutation are each associated with the serrated-adenoma pathway.[3]

The question mark in figure 1 indicates genetic or epigenetic changes specific to metastatic progression that have not been identified. Key growth factor pathways that are altered during colon neoplasia are shown at the bottom of figure 1. CIN denotes chromosomal instability, EGFR epidermal growth factor receptor, 15-PGDH 15-prostaglandin dehydrogenase, and TGF-β transforming growth factor β.[3]

An individual's risk of developing cancer depends on many factors. It is dependent on increasing age with 83% of cases arising in people who are 60 years or older.[4] Diet has also been shown to have an effect and a high intake of red and processed meat [5], as well as a high alcohol intake [5] increases the chances of developing bowel cancer.

In contrast, high fibre content in the diet has been shown to reduce the risk of colorectal cancer.[5] Other known risk factors are obesity [6], physical inactivity [7] and cigarette smoking.[8] In addition people with a first degree relative with bowel cancer are at increased risk of developing it themselves.[9]

Most cases of colorectal cancer occur in people with no family history of the disease (sporadic). Investigation of the genetic inheritance of two uncommon familial colon cancer syndromes, familial adenomatous polyposis (FAP) and hereditary non-polyposis colon cancer (HNPCC), led to dramatic breakthroughs in understanding the pathogenesis and molecular basis of the more common sporadic (non-syndromic) form of colon cancer.[2]

FAP was shown to be caused by germ line mutation of the *APC* gene (adenomatous polyposis coli). A patient with FAP carries this germ line mutation in one allele in all somatic cells, including colonocytes. This mutation underlies the development of hundreds of adenomatous polyps throughout the colon; colonic adenomas form when the second *APC* allele is damaged or lost in a colonocyte.[10–15]

Spontaneous somatic *APC* mutation in colonocytes is believed to underlie the development of sporadic adenomatous polyps. *APC* gene mutations occur early in adenoma development.[16] *APC* mutations are found in about 50% of sporadic adenomas [17] and are thought to account for 80% of sporadic colon cancers.[18] Adenomas usually remain benign. Malignant transformation requires further genetic alterations.[18]

The *DCC* (deleted in colon cancer) gene normally promotes apoptosis (cell death) and suppresses tumours. Loss of the normal DCC gene is believed to be important in the transition from an intermediate to a late adenoma.[19]

The normal *p53* gene product arrests the cell cycle following DNA injury to permit either DNA repair if the damage is correctable, or apoptosis if the damage is too severe. Mutation of the *p53* gene is believed to be important in the transition from late adenoma to carcinoma. About 50% of lesions with high-grade dysplasia and about 75% of cancers exhibit loss of normal *p53* function.[20, 22]

The *K-ras* gene encodes for a protein involved in signal transduction from the cell membrane to the nucleus.[23] Specific mutations of this gene result in constitutive activation of this signal pathway and increased colonocyte replication. These mutations are associated with exophytic growth of adenomas in the transition to carcinoma.[24] About 50% of colon cancers have *K-ras* mutations.[19]

HNPCC was shown to be caused by mutations of one of the mismatch repair genes.[24] Germ line mutations of the *hMLH1* and *hMSH2* genes account for

most of the cases. Cells with mismatch repair gene mutations cannot repair spontaneous DNA errors and progressively accumulate mutations throughout the genome with succeeding DNA replications. This progressive accumulation in oncogenes and tumour suppressor genes can result in colon cancer.[25] Mutations of the mismatch repair genes are believed to account for about 15% of sporadic colon cancers.[18]

Methylation (the addition of a methyl group) of DNA at specific sites such as at promoter regions can terminate gene expression without DNA mutation.[26] Colon cancer is sometimes associated with methylation and inactivation of the *p14* gene, normally an upstream inducer of the *p53* tumour suppressor pathway. This occurs in about 25% of colon cancers.[27] Methylation of the tumor suppressor gene *p16*, designated CDKN2A, occurs in about 35% of colon cancers.[28]

Colon cancer can occur in a pedunculated polyp, sessile polyp, mass, or stricture. Small polyps rarely contain cancer. Only about 1% of diminutive polyps contain cancer.[29] Cancer in a sessile polyp may metastasize faster than cancer in a pedunculated polyp because of the closer proximity of the lymphatic drainage.[30]

Colon cancers are classified as well-differentiated, moderately well differentiated, or poorly differentiated on the degree of preservation of normal glandular architecture and cytologic features. Progressively more poor differentiation is presumably a histological marker of further underlying genetic mutations, but the mutations associated with poor differentiation are currently unknown.[31-33]

When staging the disease the term carcinoma in situ, or high-grade dysplasia, is pathologically confined to the mucosa without penetration of the muscularis mucosa. Invasive colon cancer is commonly staged from A to C according to the Dukes system [34] (Table 1.1).

Table 1.1: Dukes classification of the stages of colorectal cancer[34]

stage A	Growth does not extend into peri rectal tissue.
stage B	Growth extends into peri rectal tissue but does not spread to lymph glands.
stage C	Growth metastasised to regional lymph nodes

Colorectal cancer is more recently staged according to the tumor–node–metastases (TNM) classification (table 1.2) [35].

Table 1.2: The 7th edition of the TNM classification of colorectal cancer by the American Joint Committee on Cancer.[35]

T Stage	Depth of primary tumour
T1	Tumour is only in the inner layer of the bowel - submucosa
T2	Tumour has grown into the muscle layer of the bowel wall – muscularis propria
T3	Tumour has grown into the outer lining of the bowel wall – into subserosa or into non-peritonealised pericolic or perirectal tissue.
T4	Tumour has grown through the outer lining of the bowel wall. It may have broken through the membrane covering the outside of the bowel (the visceral peritoneum) – T4a It may have grown into another part of the bowel, or other nearby organs or structures. – T4b
N Stage	Lymph node metastasis
N0	No regional lymph node metastasis
N1	Metastasis in 1 (N1a), 2-3 (N1b) pericolic or perirectal lymph nodes. Metastasis in the subserosa, mesentry, or pericolic-perirectal tissue without regional nodal metastasis (N1c)
N2	Metastasis in 4 or more pericolic or perirectal lymph nodes 4-6 nodes (N2a) >7 nodes (N2b)
M Stage	Distant metastasis
M0	No distant metastasis
M1	Distant metastasis (one organ M1a), (>1organ M1b)

The number system uses the TNM stages to group bowel cancers. There are 5 stages in total but stage 0 - carcinoma in situ – is non-invasive. There are 4 stages of invasive colon cancer.[35]

Table 1.3 The number system to stage colorectal cancer based on TNM [35]

Stage	
I	T1, N0, M0 or T2, N0, M0
IIa	T3, N0, M0
IIb	T4a, N0, M0
IIc	T4b, N0, M0
IIIa	T1-2, N1/N1c M0, T1,N2a,M0
IIIb	T3, N1, M0 or T4, N1, M0
IIIc	any T, N2, M0
IVa	any T, any N, M1a
IVb	any T, any N, M1b

The TNM and Dukes' classification systems correspond with each other and are used in parallel by health professionals in the menagement of colorectal cancer (Table 1.4).

Table 1.4: Correlation between TNM and Dukes classification systems[35]

TNM stage	Dukes stage
Stage I	Dukes A or B1
Stage II	Dukes B2
Stage III	Dukes C
Stage IV	-

Pathologic stage, as classified by either scheme, is highly correlated with cancer prognosis.[36]

In the diagnosis and treatment of colorectal cancer it is important that the investigations and treatment modalities patients receive adhere to certain standards and that patients have access to the most appropriate treatment for their condition. Clinical practice is about making choices. The answer to most clinical questions depends on the practitioner's knowledge, skills and attitudes, the resources available, and the patient's concerns, expectations,

and values.[37] Unfortunately there can be a variable gap between what is known from research and what is done in clinical practice. In the last 40 years attempts have been made towards a more systematic approach to the evidence that underpins medical practice.

1.2 Evidence-based medicine (EBM)

In 1972 the British epidemiologist Archie Cochrane highlighted the fact that most treatment-related decisions were not based on systematic review of clinical research. Rather they were based on ad-hoc selection of information from the vast scientific literature of variable quality, on expert opinion, or worst of all, on trial and error.[37]

Cochrane proposed that researchers and practitioners collaborate internationally to systemically review all the best clinical trials by specialty.[38] An international collaboration was indeed established and in the early 1990s funds were provided by the UK National Health Service to establish the Cohrane Centre in Oxford. This work has been continued through the Cochrane collaboration which publishes systematic reviews of randomised controlled trials (RCTs) electronically in the Cochrane Database of Systematic Reviews within the Cochrane Library. Access in many countries is available free and online.[37,39]

Also in the early 1990s, David Sacket and his colleagues from McMaster University in Ontario, Canada, coined the term ‘evidence-based medicine’ to mean ‘integrating clinical expertise and patient values with the best available external clinical evidence from systematic research to achieve the best possible patient management’.[40]

Over the years ‘evidence-based medicine’ has evolved into ‘evidence-based practice’ to include a wider provision of healthcare but the principles remain the same and that is to improve the quality of healthcare provided through systematic searching and appraisal of the research evidence.

There are several controversial areas with regard to the diagnosis and treatment of colorectal cancer. This, together with NHS funding arrangements can lead to differences in the management and availability of different treatments to patients in different areas of the UK.

Guideline development is one method by which this issue is currently being addressed and it is an attempt to ensure that the quality of the information on which healthcare decisions are based follows the principles of EBM.

1.3 Guidelines

A medical guideline is defined as a systematically developed statement to help clinicians and patients with decision-making regarding diagnosis, management, and treatment in specific areas of healthcare. The aim is to standardize medical care, to raise quality of care, to reduce risk, and to achieve the best balance between cost and effectiveness. Once a guideline is published quality standards can be developed to assist the implementation of the guideline and improve patient care nationwide.[41]

Guidelines have been in use for thousands of years during the entire history of medicine. In the late 5th century BC, Hippocrates wrote the Hippocratic oath. One of its main principles, “Do no harm” or nonmaleficence, is today the cornerstone of medical ethics. Though originally intended to guide the practice of his pupils, the oath still holds relevance today and is taken by new doctors in many countries.[42]

Guidelines are never mandatory by definition. However there are situations where by law a doctor whose practice deviates from certain guidelines may be faced with legal proceedings. Guidelines can therefore be classified into Statutory and Advisory.

1.3.1 Statutory guidelines

General Medical Council guidance

Statutory guidelines are subject to the Law and for doctors this type of guidance is produced by the General Medical Council (GMC).

The purpose of the GMC is to protect, promote and maintain the health and safety of the public by ensuring proper standards in the practice of medicine. The GMC was established under the Medical Act of 1858. Over time a range of new legislation has been introduced that defines its powers and responsibilities. To practice medicine in the UK all doctors are required by law to be both registered with the GMC and hold a license to practice. Licensed doctors are required to demonstrate to the GMC that they are practising in accordance with the generic standards of practice set by the GMC.[43]

Good Medical Practice (2006)[44] is the core guidance which the GMC produces for doctors regarding their fitness to practice. This guidance sets out the principles and values on which good practice and medical professionalism is founded. Quoting directly from the Good Medical Practice Document 2006 "It is the responsibility of every doctor registered with the GMC to be familiar with Good Medical Practice (2006) and to follow the guidance it contains. It is guidance, not a statutory code, so every doctor must use their judgement to apply the principles to the various situations they are faced with. Every doctor must be prepared to explain and justify his or her decisions and actions. "Serious or persistent failure to follow this guidance will put your registration at risk."[44]

This is therefore a situation where a by law a regulatory body has the power to remove a doctor's licence to practice if the doctor fails to justify appropriately the reasons from deviating from the set guidance.

1.3.2 Advisory Guidelines

Although the production of advisory guidelines does not directly relate to an act of parliament in the same way as the GMC guidelines do, failure by a doctor to comply with advisory guidelines can also lead to dispute with an employer or in the case of patient harm or percieved harm to negligence proceedings. It is important to highlight that “a doctor is not guilty of negligence if he has acted in accordance with a practice accepted as proper by a responsible body of medical men skilled in the relevant art.” (Bolam test 1957).[45] Therefore once again the doctor who goes against accepted professional guidance needs to be able to justify his actions appropriately. There are a variety of bodies that produce advisory guidelines.

Guidance produced by Government (e.g. Department of Health (DoH), Welsh Assembly, Scottish Parliament)
These are usually service related guidelines (e.g. referral pathways, referral timelines). They are usually consensus statements that are the result of working groups where invited specialist(s) have been asked to give their opinion(s) to policy makers. Adherence is often strongly recommended by employment contracts with health providers or government agencies and it may be a contractual obligation to abide. There may be financial incentives or penalties which are often used to ensure compliance.

Guidance produced by local health-care providers (e.g. healthcare trusts or local networks)
These are both service and clinical guidelines. Typical examples are local prophylactic antibiotic prescription guidance, thrombosis prophylaxis guidance or blood transfusion guidance. These may be evidence-based or based on other published guidelines (from national guideline developers or professional bodies such as professional colleges / associations) with appropriate adaptation to the local community being served. They may also be consensus statements from local committees. These committees are usually made up of relevant specialist staff that volunteer their time to represent their department or specialty on the committee. Internationally there is a recent

trend for local hospitals to employ professional guideline developers to oversee their guideline and protocol development (Australia, USA). Adherence is strongly recommended and certain employment contracts include a clause regarding possible legal action (under employment law) against the employee who fails to adhere to local protocols or agreed processes.

Guidelines produced by The National Institute of Health and Clinical Excellence (NICE).

NICE is an independent body commissioned in 1999 by the Department of Health to produce guidelines for healthcare professionals treating patients in the NHS in England and Wales. NICE guidelines are evidence-based recommendations designed to promote good health and prevent ill health. The guidelines address both clinical-effectiveness and cost-effectiveness issues.[46] There are different types of NICE guidelines.

Clinical guidelines

Clinical guidelines cover aspects of the management of a particular disease or condition. The evidence supporting different treatments is examined to assess whether they are effective for patients. The guidelines make recommendations on which treatments should be made available in the NHS in England and Wales, in order to ensure the best care is available to all patients. Clinical guidelines sit alongside, but do not replace the knowledge and skills of experienced health professionals and consider both the clinical effectiveness and also the cost effectiveness of cancer treatments.

Service guidelines

Service guidelines make recommendations on how NHS services for patients should be organised in England and Wales. Both the anticipated benefits and the resource implications of implementing the recommendations are considered.

Technology Appraisal guidance

Technology appraisal guidance focuses on the clinical and cost effectiveness of one or more technologies, such as new drugs, surgical procedures and medical devices.

Interventional Procedure Guidance

Interventional procedures (IP) guidance covers the safety and efficacy of interventional procedures used for diagnosis or treatment.

Public Health Guidance

Public health guidance deals with promoting good health and preventing ill health.

Guidelines produced by The Scottish Intercollegiate Guidelines Network (SIGN)

SIGN develops evidence based clinical practice guidelines for the NHS in Scotland. SIGN guidelines are derived from a systematic review of the scientific literature. SIGN guidelines are produced by guideline development group members, with support from the SIGN Executive according to structured robust methodology.[46,47]

Guidelines produced by The Guidelines Audit and Implementation Network (GAIN.)

GAIN produces guidelines for the NHS in Northern Ireland. Its role is safety and quality improvement in Health & Social Care Services throughout Northern Ireland through the commissioning of regional audit and guidelines as well as the promotion of good practice through the dissemination of audit results, and the publication and facilitation of implementation of regional guidelines.[48]

Guidelines produced by international guideline developing bodies

Since the establishment of NICE in the UK other countries are also establishing guideline-developing bodies. In the US there is the Agency for

Healthcare Research and Quality.[49] In The Netherlands, two bodies (CBO and NHG) publish specialist and primary care guidelines, respectively.[50] The German Agency for Quality in Medicine (ÄZQ) coordinates a national program for disease management guidelines [51, 52]. These organisations are members of the Guidelines International Network (G-I-N). [50] G-I-N is owner of the International Guideline Library – the largest web-based database of medical guidelines worldwide - and pursues a set of activities aiming at promoting best practice and reducing duplication in the guideline world. The USA and other countries also maintain medical guideline clearinghouses. In the USA, the National Guideline Clearinghouse (NGC) maintains a catalogue of high-quality guidelines published by various organizations. In addition, the National Comprehensive Cancer Network, an alliance of leading US cancer centres also provided reputable guidelines.

Guidelines produced by professional medical organisations and societies

Specialist working groups formed by members of the executive who have volunteered to sit on the guideline panel usually produce these guidelines. An alternative is that panel members are self-selected or nominated by their peers. The evidence provided varies with some societies producing very good quality evidence-based guidelines and others producing a higher number of consensus statements particularly when addressing topics where evidence is not available in the literature.

1.4 Health Economics

Health economics is concerned with issues related to efficiency, effectiveness, value and behaviour in the production and consumption of health and health care.[52] It is about improving the health of a population through the efficient use of resources.[53]

No country can afford all the health care interventions that might benefit patients. Clinical need will always outstrip available resources so priorities have to be agreed. How this prioritisation process takes place varies from

country to country but the need to prioritise in some way is clear. There just is not (and never will be) enough money to provide every possible service.[54] Health economics applies at all levels, including individual clinical decisions. Clinicians already take resources and value for money into account in clinical decisions, and the incorporation of good-quality health-economic evidence into clinical guidelines can help make this less arbitrary and more consistent.[53]

In Britain, before NICE was established, these decisions were largely made behind closed doors. Although formal economic assessments were sometimes made they were rarely exposed (or explained) to the public, nor was the public involved in making the assessments.[54]

More often, decisions about how NHS money was used were based on other factors. These included historical patterns of health care, assumptions about where (and how) additional investment might appropriately be made, pressure from special interest groups, political lobbying and perceptions about public preferences.[54]

The creation of NICE made possible a fundamental change in how these issues were tackled by the NHS. For the first time, a national public body was charged with making authoritative recommendations about the availability of new and established treatments, and pathways of care, and doing so formally taking cost-effectiveness (or value-for-money) into account.[54]

Few healthcare systems had tried to do this before. In those countries that did, such as in Australia, the process was mainly limited to new pharmaceutical products.[54]

NICE guidelines assess the clinical effectiveness and cost effectiveness of treatments and ways of managing a particular condition. Cost effectiveness is the estimated costs of the treatment options in relation to their expected health benefits rather than the total cost or resource impact of implementing them.[54]

In evaluating healthcare and making its decisions on whether an intervention should be available to the NHS, NICE compares interventions by using an economic approach called ‘cost-utility analysis’. This considers the impact each intervention has on health compared to current care and how much it costs again compared to the costs of current care.[54]

NICE health economics carry out the cost-effectiveness analyses with the units of effectiveness expressed in QALYs (Quality Adjusted Life Years). QALYs are an overall measure of health outcome that weighs the life expectancy of a patient with an estimate of their health-related Quality of Life (QoL) measured on a 0–1 scale.[54]

The QALY captures the treatment on both ‘quality of life (QoL)’ and length of life. One QALY is the equivalent of one year in perfect health, or two years in 50% of that health, or four years in 25% of that health, and so on. It provides a ‘common currency’ that allows different interventions to be compared for different conditions.[54]

The use of QALYs is widely recognised as a useful approach for measuring and comparing the efficiency of different health interventions. There are however well-documented methodological problems with QALYs, but this is also true of other approaches. If there is insufficient data to estimate QALYs gained, an alternative measure of effectiveness may be considered for the cost-effectiveness analysis (such as life years gained or cases averted, or a more disease-specific outcome).[54]

The cost per QALY indicates how much extra it costs the NHS to buy the equivalent of one QALY of benefit from a new intervention over and above what it pays now for the benefits from existing treatments.[54]

Economists also refer to this comparative ‘cost-per-QUALY’ as the ‘incremental cost effectiveness ratio (ICER)’. The incremental cost-effectiveness ratio (ICER) is an equation used commonly in health economics

to provide a practical approach to decision making regarding health interventions. It is typically used in cost-effectiveness analysis.[55] ICER is the ratio of the change in costs to incremental benefits of a therapeutic intervention or treatment.[56]

The equation for ICER is: ICER = (C1 – C2) / (E1 – E2) where C1 and E1 are the cost and effect in the intervention or treatment group and where C2 and E2 are the cost and effect in the control care group.[55] Costs are usually described in monetary units while benefits/effect in health status is measured in terms of quality-adjusted life years (QALYs) gained or lost.[55]

ICER provides a means of comparing projects or interventions across various disease states and treatments. As seen in the equation above, a ratio is created with the units of cost per benefits/effect unit. By using this ratio, comparisons can be made between treatment modalities to determine which provides a more cost-effective therapy.[55]

For example, when one strategy is more effective but also more costly, then the magnitude of the incremental cost-effectiveness ratio (ICER) should be considered. The cost per QALY gained is calculated as the difference in mean cost divided by the difference in mean QALYs for one strategy compared with the next most effective alternative strategy. If one intervention appears to be more effective than another, the guideline development group will have to decide whether the increase in cost associated with the increase in effectiveness represents reasonable ‘value for money’.[54]

Some people feel ICER studies provide an opportunity to help contain health care costs without adverse health consequences.[57] They also provide to policy makers information on where resources should be allocated when they are limited.[55] As health care costs have continued to rise, many new clinical trials are attempting to integrate ICER into results to provide more evidence of potential benefit.[58]

Others feel that basing health care interventions on cost-effectiveness is a type of health care rationing and have expressed concern that using ICER will limit the amount or types of treatments and interventions available to patients.[55]

The aim of the cost-utility approach used by NICE is to use the budget of the NHS to ‘purchase’ the greatest number of QALYs possible i.e. to maximise the amount of health gained for the money available. [54]

NICE has never identified an ICER above which interventions should not be recommended and below which they should.[60]

However an ICER threshold range has been set (£20-30,000 per QALY) to indicate a point at which factors other than the ICER itself should be examined and debated as part of the judgment about the acceptability of the intervention as an effective use of NHS resources.[59,60]

Such factors include: the degree of certainty around the ICER in the cost-effectiveness analysis, a change in the quality of life inadequately captured in the representation of the health gain, and demonstrable benefits inadequately captured in the measurement of health gain.[59,60]

The overall ‘rationing of healthcare’ debate, the use of QALY’s, and the ICER threshold debate remain and will probably remain at the forefront of healthcare related debates; however this is the situation in the UK at present.

1.5 The guideline debate

Opinion on the value of guidelines differs amongst physicians. Advocates of guidelines believe that they are a welcome development that brings improvement to clinical practice. They believe that finding, evaluating and implementing the results of medical research can, and often does, make patient care more objective, more logical, and more cost effective. [60]

They maintain that the purpose of guidelines is to improve quality of healthcare provision, to make evidence-based standards explicit and

accessible, to make decision-making in the clinic and at the bedside easier and more objective, to educate patients and professionals about current best practice, and to improve the cost effectiveness of health services.[61]

There are those however, who see guidelines as a potential danger. They fear that when guidelines are applied in a vacuum (that is in the absence of common sense and without regard to the individual circumstances and priorities of the person being offered treatment), the evidence based approach to patient care is a reductionist process with a real potential for harm, particularly when becoming the only accepted option for example in the context of a busy practice or a resource-strapped health-care provider. [60]

Others just feel overwhelmed by the sheer number of guidelines available, finding it often conflicting or confusing. There are also those that worry that guidelines are an unnecessary external authority that removes autonomy from the individual clinician.[61]

Some take the view that guidelines developed at national or regional level may not reflect local needs or have the ownership of local practitioners. Guidelines developed in or for secondary care may not reflect demographic, clinical or practical differences between this sector and the primary care setting.[61]

Guidelines may produce undesirable shifts in the balance of power between different professional groups (for example between clinicians and academics or purchasers and providers) and guideline development may be perceived as a political act.[61]

Others see it as an undesirable way to provide a yardstick for assessing professional performance, to delineate the division of labour (e.g. between GPs and hospital consultants), or even as a tool for external control.[61]

Evidence-based medicine and guidelines are now so common in clinical practice that it is hard to remember a time before them.[62] Yet there are

those who question the validity of the development process and the quality of the evidence on which guidelines are often based.[61,62, 63]

Doctors who are sceptical about the scientific basis of guidelines have two choices: they can follow guidelines even though they suspect doing so will cause harm, or they can ignore them and do what they believe is right for their patients, thereby risking professional censure and possibly jeopardising their careers.[63]

This is no mere theoretical dilemma. There is evidence that even when doctors believe a guideline to be harmful and compromised by bias, a substantial number follow it.[63]

It is important therefore for research in the field of guideline development to take place. In-depth analysis of the methodology and the body of evidence that support the recommendations will ensure thorough understanding of the process, its strengths and weaknesses. It will provide an opportunity for quality assessment and the platform for any necessary improvements.

2.0 Aims

To research the evidence base on which the NICE guideline for the diagnosis and management of colorectal cancer is developed. This principle aim has been broken down into the following six study aims.

1. To research the availability and quality of guideline methodology data that can inform all end users of guidelines produced by NICE on the guideline development process for the purpose of guideline appraisal (Chapter 4).

2. To research the availability and quality of epidemiology data that can inform a guideline development group (GDG) on incidence, mortality, survival and prevalence of CRC for the purpose of the CRC guideline needs assessment report (Chapter 5).

3. To research the availability and quality of current clinical practice data that can inform a GDG on aspects of CRC management for the purpose of the CRC guideline needs assessment report (Chapter 6).

4. To research the availability and quality of the diagnostic data that make up the evidence to support the most effective method for diagnosing liver metastases from CRC to assess resectability (Chapter 7).

5. To research the availability and quality of the therapeutic data that make up the evidence to support the most effective method for the follow up of patients that have been diagnosed and treated for primary CRC (Chapter 8).

6. To research the international nature of data supporting the NICE guideline on CRC and the influence this may have on the resulting recommendations (Chapter 9).

3.0 Methodology and Methods

3.1 Research Methodology

In 2009 NICE commissioned the National Collaborating Centre for Cancer (NCC-C) to develop a clinical guideline on the diagnosis and management of CRC. This marked the beginning of a two-year project that culminated in the publication of the guideline in 2011.

For the first time in parallel to the guideline development process the NCC-C introduced the opportunity for research into guideline methodology and for this reason granted full access to the entire guideline development process of the NICE CRC guideline.

The research was carried out through active membership of both the NCC-C guideline development team and the NICE guideline development group (GDG) of health professionals established to produce the recommendations. Though participation in all aspects of the guideline development process were encouraged the researcher held a non-voting role during the final formulation of the recommendations similar to the members of the NCC-C guideline development team.

This thesis is the result of the research endeavour described above. Six thematically distinct studies were designed prior to the commencement of the guideline development process. These all attempt to answer the principle aim, to research the evidence base on which the NICE CRC guideline is developed. The decision was taken to design multiple distinct studies so as to cover as many different aspects of the guideline development process as possible within the given time limitations of the project. Each study answers each of the six secondary aims and is presented in separate chapters.

The detailed methods used to carry out each individual research study are presented within each study chapter. A summary of these methods is presented below.

3.2 Summary of research study methods

Methods employed to address aims 1, 2 & 3

Online data mining.

Web based research was carried out with the purpose to identify the data that was relevant to answer each research aim. The details of the search engine, the search strategy, the search terms used, are included in the methods section of each of the relevant chapters (4,5,6).

Methods employed to address aim 3

Data request from national databases of patient information

National databases of patient information were approached and anonymised data on current clinical practice was requested. Linkage studies combining data from multiple databases were also discussed and planned. Details of the search strategies are included in the methods section of chapter 6.

Methods employed to address aims 4 & 5

Systematic reviewing according to NICE methodology

The review questions were broken down into appropriate search terms and the appropriate search strategies were created. The international medical literature was systematically searched through multiple databases and registers.

Study selection was carried out based on pre-determined inclusion and exclusion criteria.

Study quality assessment was carried out using appropriate quality checklists for each type of study under consideration. The checklists were all validated tools used within the international guideline community and approved by NICE for use as part of their guideline development.

The data was extracted and analysed. Where data synthesis was appropriate meta-analysis software was used (REVMAN 5).

Methods employed to address aim 6

Review of 'Linking Evidence To Recommendation' (LETR) data

Information on how the evidence that supported the guideline was linked to the final recommendations formulated by the guideline development group was collected through:

i) participation in all the guideline development group (GDG) meetings throughout the two year development of the CRC guideline and experiencing in close proximity the process of developing recommendations from the body of evidence presented for each topic.

ii) reviewing of all the final formal LETR sections of the CRC guideline document for each of the topics the guideline addressed for evidence of the recommendations having been influenced by the national setting of the evidence that supported that recommendation.

4.0 The availability and quality of methodology data for guideline appraisal

4.1 Introduction

Methodology is the systematic analysis of the methods applied to a field of study.[64]

It offers the theoretical underpinning for understanding how a specific piece of research has been carried out. Many treat it as a synonym for method or body of methods.[64]

Doing this shifts it away from its true epistemological meaning and reduces it to being the procedure itself, the set of tools or the instruments that should have been its outcome. [64]

A methodology is the design process for carrying out research or the development of a procedure and is not in itself the instrument.[64]

For Guidelines, there is a standardised way of appraisal and quality assessment of methodology by the use of the AGREE tool.[65]

The Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument (appendix 1) evaluates the process of practice guideline development.[65]

It is a tool that assesses the methodological rigour and transparency in which a guideline is developed and it is used internationally.[65]

The original AGREE Instrument has been updated and methodologically refined. [66]

AGREE II is not only used for appraisal but can also be used to provide a methodological strategy for the development of guidelines.[66]

In addition, it can be used to guide what information and how that information ought to be reported in guidelines.[66]

AGREE II is designed for guidelines developed by local, regional, national or international groups or affiliated governmental organizations.[67]

Amongst the medical profession there are those who question the reasoning behind a guideline developing organisation such as NICE.[61]

Some worry that guidelines produced by non-professional bodies may be intellectually suspect and a means of external control, policing, and removal of autonomy from the professional medical groups. [61]

There is a fear amongst some of the professionals that such guidelines might be a way in which politicians and health-service managers who have made use of the rationale for evidence-based guidelines will use them to make judgements that serve their own political or economic agendas.[61]

In this way they would be using evidence that proves interventions are effective “on average” but which omit the value of experience and professional insight.[61]

Guidelines have been produced for many years by the Royal Colleges and specialist professional societies.[68]

As the field of evidence based medicine has been evolving guideline development by such organisations has in the past been characterised by problems that potentially undermine the quality and trustworthiness of the guidelines they produce.[68]

This includes lack of transparency, limitations in the process of systematic reviews, but more importantly a failure to use rigorous methodology.[68]

The rapid proliferation of guideline production often means guidelines are produced in parallel either by professional associations with an overlap in interest or by professional associations and national guideline developing bodies.[68]

This, apart from appearing unecessary and wasteful of resources, can potentially lead to confusion amongst the audience when guidelines are produced by different bodies on the same topic with differences in their recommendations.[68]

In the UK National Health Service, all doctors, nurses, pharmacists, and other health professionals now have a contractual duty to provide clinical care based on best available research evidence. [61]

Furthermore, whilst the medico-legal implications of “official” guidelines have rarely been tested in the UK, US courts have ruled that guideline developers can be held liable for faulty guidelines.[61]

In addition, the same ruling states that doctors cannot pass off their liability for poor clinical performance by claiming that adherence to guidelines corrupted their judgment.[61]

The ability to appraise the quality of a guideline irrespective of the authority that has produced it is vital.

In order to do so it is important to be able to access and assess the quality of the guideline development process by which it was produced. This is done by appraisal of the methodology.

4.2 Aim

To research the availability and quality of guideline methodology data that can inform all end users of guidelines produced by NICE on the guideline development process for the purpose of guideline appraisal.

4.3 Methods

Reasearch on the availability of NICE methodology data was carried out by online data mining of the NICE website according to a predetermined search strategy as outlined below.

The methodology data retrieved was also reviewed for information regarding the quality strategy of the development process. It was specifically scanned for information on whether a standardised tool like the AGREE II, or a different tool or method, was used.

The NICE methodology data retrieved was also further assessed for its availability and quality by comparison with similarly retreaved methodology data from other UK guideline developing bodies.

Online Data mining

NICE methodology data

The search engine ‘google’ was used to search for available data regarding the NICE methodology. The search terms that were used were: ‘NICE’, ‘guideline’, ‘methodology’, ‘development’, ‘process’

The information produced from the search was selected if it answered the following questions:

Is this information about the NICE clinical guideline development process?

Is this information about the quality assessment of the guideline process with the AGREE tool or any other mentioned guideline quality appraisal/development tool?

Is this information about quality assessment of the development process without a standardised instrument?

An overall assessment was finally made by the reviewer regarding the ease of access to the data through the online data mining process.

Methodology data from other UK guideline producing bodies

A systematic search was performed of the information presented on the websites of all current UK surgical societies, associations, and Royal Colleges.

The list of surgical societies to be searched was drawn up by associating a society for each of the known surgical specialties and sub-specialties. One medical society was included on the list as it was felt that this medical specialty through endoscopy practice overlapped with surgical practice and should therefore be included.

Search terms relating to the associations' titles were entered into the 'google' search engine and the precise title and webpage address of the society was identified. In the case that surgical subspecialties were found to have more than one associated society, all were included in the search.

Table 4.1: List of professional societies and Royal Colleges included in the search for methodology data.

Organisation	**Web address**
Royal College of Surgeons of England	www.rcseng.ac.uk
Royal College of Surgeons of Ireland	www.rcsi.ie
Royal College of Physicians and Surgeons of Glasgow	www.rcps.ac.uk
Royal College of Surgeons of Edinburgh	www.rcsed.ac.uk
Association of Surgeons of Great Britain and Ireland ASGBI	www.asgbi.org.uk
Association of Coloproctology of Great Britain and Ireland ACPGBI	www.acpgbi.org.uk
Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland AUGIS	www.augis.org

The Vascular Society of Great Britain and Ireland	www.vascularsociety.org.uk
Association of Breast Surgery UK - ABS at BASO	www.baso.org
British Association of Surgical Oncology BASO	www.baso.org
Association of Laparoscopic Surgeons of Great Britain and Ireland	www.alsgbi.org
British Association of Day Surgery BADS	daysurgeryuk.net
British Hernia Society	www.britishherniasociety.org
British Orthopaedic Association	www.boa.ac.uk
British Trauma Society	www.bts-org.co.uk www.trauma.org
British Association of Otorhinolaryngologists ENT-UK	www.entuk.org
The British Association of Urological Surgeons	www.baus.org.uk
British Association of Plastic, Reconstructive and Aesthetic Surgeons BAPRAS	www.bapras.org.uk
British Association of Aesthetic Plastic Surgeons BAAPS	baaps.org.uk
Society for Cardiothoracic Surgery of Great Britain and Ireland	www.scts.org
British Society of Gastroenterology	www.bsg.org.uk

Each website was visited and carefully searched for data on guidelines and guideline methodology. The home webpage was searched as a first step followed by all linked webpages accessible through the home page option lists. All lists and linked pages were searched for data, not just those that immediately appeared relevant to guideline data such as lists referring to ‘research’ or ‘resources’.

If no information on guideline or guideline methodology was available through this route then the home webpage ‘search’ option was used and the key words ‘guideline’ and ‘methodology’ entered as separate searches not as a limited combined search.

If a society website was found to offer guideline documents but had no information available on the website about the guideline methodology then all guideline documents were opened and their contents page searched for a methodology section. Paper versions of guideline documents were not requested and societies were not contacted via telephone or mail for methodology data.No formal statistical software was necessary in the analysis of the results as numbers involved were small enough to analyse manually.

4.4 Results

NICE methodology data

Availability of methodology data

The methodology and all related guideline methods and tools are publicly available and easily accessible online through the NICE website and the webpage of the guideline development manual.[69]

The NICE guideline development process [69]

Referral and remit

The Department of Health asks NICE to produce a guideline on a particular topic. The topics for guidelines are based on recommendations from topic selection consideration panels.The topic referral is also associated with a remit that identifies the broad areas to be covered.

National Collaborating Centres (NCC)

NICE commissions one of the four National Collaborating Centres (NCC) to co-ordinate the development of a guideline. For the clinical guideline on the management of colorectal cancer this was the NCC for cancer (NCC-C).

This is responsible for developing NICE guidelines for the NHS in England and Wales on treating and caring for people with cancer. The other national collaborating centres are the centre for acute and chronic disease, the centre for women and children, and the centre for mental health.

A management board comprising representatives of relevant professional bodies oversees the work at the NCC-C. This currently includes representatives for a number of Royal

Colleges, the National Cancer Research Institute (NCRI), and other charity and academic bodies.

The management board meets regularly and among its many functions is to oversee the guideline development process for each guideline.It advises the NCC-C on negotiating with NICE on quality issues.

The guidelines team in the Centre for Clinical Practice at NICE supports and advises the NCC during the process.

Guideline Development Group (GDG)

A Guideline Development Group (GDG) is established to manage the work. The GDG is composed of health professionals who are involved in the treatment and management of patients with cancer.

It also includes at least two patient/carer representatives. GDGs usually consist of 12-15 people. NICE is not represented on the GDG.

The scope

The remit is translated into the scope. The scope provides a framework within which to conduct the guideline development work. When developing the scope key clinical issues are selected by the scoping group. This group consists of representatives of the GDG, NCC-C technical team and NICE.

Stakeholders

Before the scope consultation takes place stakeholders are invited to a scoping workshop to discuss the key clinical issues identified by the scoping group. Stakeholder organisations are organisations with an interest in a particular guideline. They

register with NICE at the beginning of the guideline development process and contribute their views during consultation periods.

In the NICE clinical guideline development process, stakeholders are: national patient and carer organisations, national organisations that represent healthcare professionals, companies that manufacture medicines or devices used in the clinical area covered by the guideline, providers and commissioners of health services in England and Wales, statutory organisations including the Department of Health and research organisations.

After a consultation period the scope is finalised. The scope provides information to healthcare professionals, stakeholders and the public about the expected content of the guideline.

Review questions

The key clinical issues listed in the scope are next broken down into review questions. The exact number of review questions for each clinical guideline depends on the topic and the breadth of the scope. However, the number of review questions must be manageable for the GDG and the NCC technical team within the agreed timescale.

For standard clinical guidelines that take 10–18 months to develop (from the time the scope is signed off to submission of the draft guideline), between 15 and 20 review questions is a reasonable number. This number is based on the estimate that, on average, it is feasible for a maximum of two systematic reviews to be presented at any one GDG meeting.

Review questions are usually drafted by the NCC technical team. They are then refined and agreed by all GDG members through discussions at GDG meetings.

The different perspectives among GDG members will help to ensure that the right review questions are identified, thus enabling the literature search to be planned efficiently.

Often the main questions need refining again once the evidence has been searched, and this may generate sub-questions.

Review questions for economic analysis

Questions are selected for economic analysis as a joint decision between the health economist and the other GDG members. The health economist is a core member of the GDG alongside the rest of the NCC technical team, and is involved at the earliest opportunity and attends all GDG meetings.

The expertise of all of the GDG members is necessary to ensure that economic evidence is underpinned by the most plausible assumptions and the best available clinical evidence. Selection is based on potential value across all key clinical issues, quality of available evidence, and time available for economic modeling.

There are likely to be large differences between clinical guideline topics in the amount, relevance and quality of the economic literature. In some topic areas there may be high-quality data that can be used in economic models, whereas in other areas there will be little information.

Defining the economic priorities for each clinical guideline starts during scoping, and proceed alongside development of the review questions.

Systematic reviewing

A systematic review is carried out for each question. Review questions are broken down into different parts and used to

devise a search strategy using the PICO (patient, intervention, comparison and outcome) framework.

This can be constructed from terms relating to the population, combined with terms relating to the interventions and comparisons to be evaluated.

The search strategy is discussed and approved by the GDG. The searches are then carried out by the information specialist. Core and subject-specific databases are searched. Other sources such as registers are also included.

Before acquiring papers for assessment, the systematic reviewer sifts the evidence identified in the search in order to discard irrelevant material. Next, the remaining abstracts are scrutinised against the inclusion criteria agreed by the GDG.

Abstracts that do not meet the criteria are excluded. Any doubts about inclusion should be resolved by discussion with the GDG before the results of the study are considered.

Once the sifting is complete, full versions of the selected studies are acquired for assessment. Studies that fail to meet the inclusion criteria once the full version has been checked are excluded; those that meet the criteria are assessed.

Because there is always a potential for error and bias in selecting the evidence, double sifting (that is, sifting by two people) of a random selection of abstracts is performed periodically.

Once a study has been selected it is assessed using a methodology quality checklist. Data is extracted to a standard template for inclusion in an evidence table.

Meta-analysis may be needed to pool treatment estimates from different studies.

Developing recommendations

In developing recommendations the GDG must decide what the evidence means in the context of the review questions and economic questions posed.

There are many reasons why it can be difficult for a GDG to reach a decision about a recommendation. The literature search may have found no evidence. The quality of the evidence may be poor. There may be conflicting evidence. The clinical evidence may not be directly applicable to the population covered by the guideline.

The GDG may have to consider consensus methods to identify best practice. The reasoning behind all decisions are documented and presented in the full guideline.

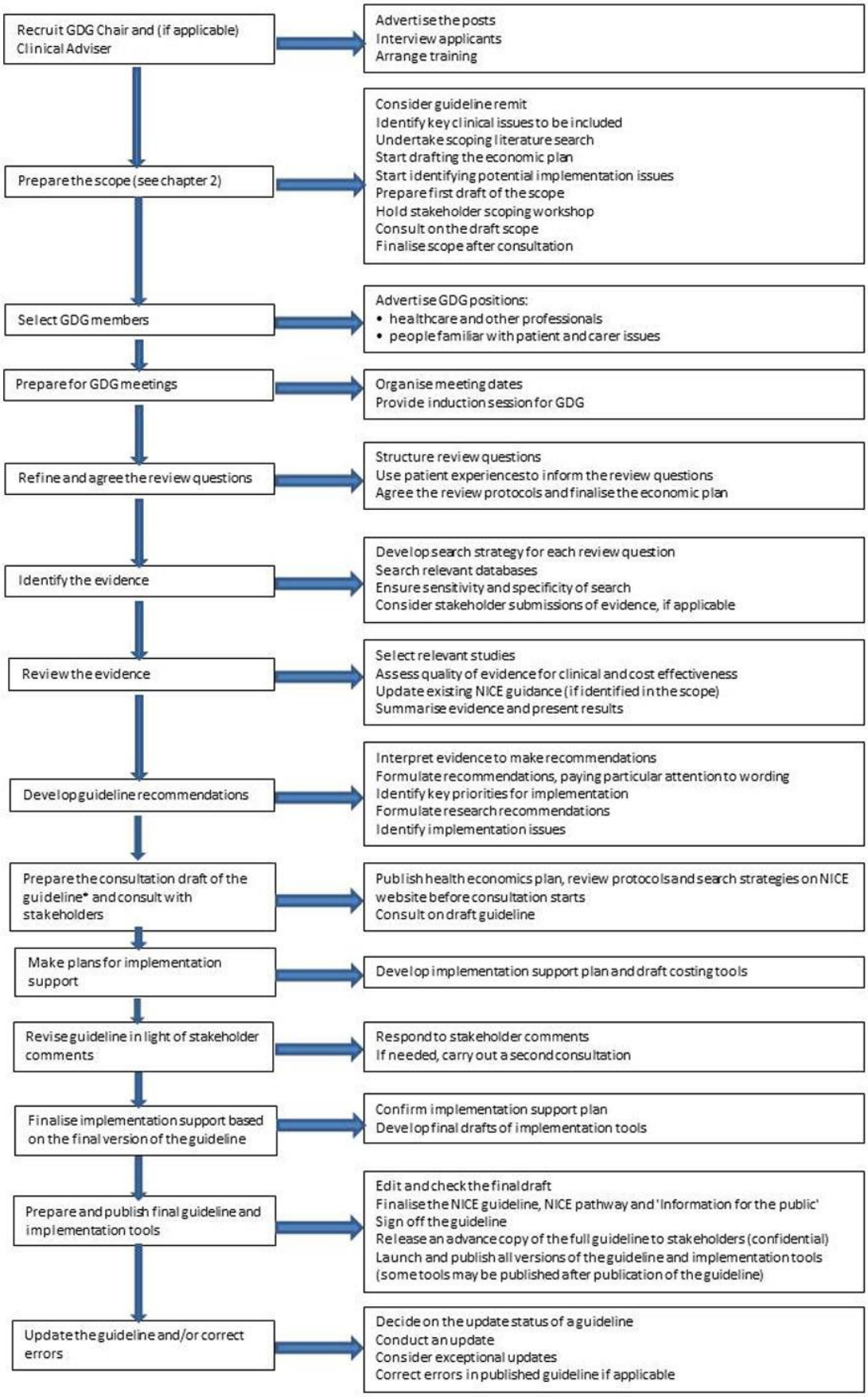
Final consultation and publication

At the end of the process, and after further consultation with stakeholders, NICE's Guidance Executive signs off the guideline.

The Guidance Executive confirms that the NCC has developed the guideline in accordance with the terms of the remit from the Secretary of State for Health and the scope, and by following NICE's process and methods.

The guideline is then published and distributed to the NHS in England and Wales.

Figure 4.1: Summary of NICE clinical guideline development process [70]



Quality of methodology data

The following information was identified in the NICE methodology manual on the NICE website relating to guideline methodology quality assessment:

> "NICE methodology follows the AGREE II framework. NICE guideline development aims to be a transparent process using the principles of evidence-based medicine." [46]

> "The NICE guideline development process has been drawn on the advice of international guideline development methodology experts, internationally acceptable criteria of quality, the expertise of the clinical guidelines team in the Centre for Clinical Practice (CCP) at NICE, and the experience of the staff at the national collaborating centres (NCCs) where the guidelines are produced on behalf of NICE."[69]

Methodology data from other UK guideline developing bodies

Availability and quality of methodology data

Table 4.2: Availability and quality of methodology data of UK surgical societies and Royal Colleges.

Society	Produce guidelines	Guideline methodology available online or in guideline document	Guideline development quality checklists used e.g. AGREE tool
Royal College of Surgeons of England	Yes –more audit from the clinical effectiveness unit	No	No
Royal College of Surgeons of Ireland	Yes – 1st under development	No	No
Royal College of Surgeons of Glasgow	No – they provide links to SIGN	No	No

Royal College of Surgeons of Edinburgh	No – they provide links to SIGN	No	No
Association of Surgeons of Great Britain and Ireland ASGBI	Yes	No	No
Association of Coloproctology of Great Britain and Ireland ACPGBI	Yes	No	No
Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland AUGIS	Yes – but also refer to NICE and SIGN	No	No
The Vascular Society of Great Britain and Ireland	Yes	No	No
Association of Breast Surgery UK ABS at BASO	Yes	No	No
British Association of Surgical Oncology BASO	Yes – just breast	No	No
Association of Laparoscopic Surgeons of Great Britain and Ireland ALS-GBI	Yes	No	No
British Association of Day Surgery BADS	Yes	No	No
British Hernia Society	No	NA	NA
British Orthopaedic Association	Yes	No	No
British Trauma Society	No access to website without membership	NA	NA
British Association of Otorhinolaryngologists ENT-UK	Yes	No	No
The British Association of Urological Surgeons	Yes	No	No
British Association of Plastic, Reconstructive	No	NA	NA

and Aesthetic Surgeons BAPRAS			
British Association of Aesthetic Plastic Surgeons BAAPS	No	NA	NA
Society for Cardiothoracic Surgery of Great Britain and Ireland	No access to website without membership	NA	NA
British Society of Gastroenterology	Yes	Yes	Yes

- 21 website pages were visited and of those 19 allowed access without membership. Of those, 18 were websites of surgical societies and Royal Colleges and 1 was of a medical society of relevance to CRC
- Of the 18 surgical society websites 14 produced guidelines for their members (77.8%)
- No surgical society website provided information online about their guideline methodology
- Of all 21 societies that produced guidelines only 1 society website provided information about their guideline methodology (5.2%). It was this same organisation that reported on their website about using the AGREE tool as a template for their guideline development and reporting

4.5 Discussion

The results show that NICE provides easily accessible information on the methodology of its guideline production. The guideline manual is an extensive and detailed document easily accessible online and clearly describing all processes relating to NICE guideline development.

Guideline development at NICE has been set up after consultation with a variety of methodology and guideline experts and follows the AGREE tool in its development strategy, an international tool for guideline quality assessment and development.

The results however show very poor online reporting of guideline methodology by all surgical societies and Royal Colleges.

Even though the majority of the organisations produce guidelines which are available online for their members and the public, they do not supply data on the methodology of these guidelines online.

This does not mean that the guidelines produced are of poor quality but it does leave an uncertainty about their quality.

The data is limited to societies relevant to surgery and may not be indicative of availability of methodology relating to guidelines produced by medical societies or indicative of the quality of the reporting of surgical professional bodies internationally.

The results may also be a reflection of the fast pace of information technology advancement. The majority of these societies may be unable to keep up with accurate representation of the guideline data they provide via the technical medium of the internet.

In addition, the pace of change and evolution is fast within the discipline of evidence based medicine. The importance of principles and details of guideline development are being highlighted with the passage of time and more importantly are being registered by the guideline community with time and with maturation of this entire field of research.

The acknowledgment of methodology developments relies heavily on enthusiasts keeping up to date and applying these improvements to the guideline work they are doing.

The data however does highlight an important gap in guideline presentation by surgical societies in the UK. The absence of methodology data from online guideline documents is serious.

With no methodology to refer to and to appraise it is difficult or impossible to draw conclusions about:

i) The overall purpose of the guideline: it is important to be able to define the health question addressed and the target population.

ii) The extent to which the guideline was developed by the appropriate stakeholders and to which it represents the views of its intended users.

It is important that the guideline development group includes individuals from all the relevant professional groups, with an appropriate level of expertise and that the process of their selection is transparent.

It is also important that the views and preferences of the target population have been sought. The target users of the guideline also need to be clearly defined.

iii) The process used to gather and synthesise the evidence, the search dates, the search engines and databases employed, and the methods used to grade the evidence.

It is important to be able to assess whether systematic methods were used to search for the evidence, that the criteria for selecting the evidence were clearly set out at the beginning and that the strengths and limitations of the body of evidence are clearly described.

It is also important to be able to identify clearly the methods used to formulate the recommendations, and that there is an explicit link between the recommendations and the supporting evidence.

Additionally it is important that the side effects and risks have been considered in formulating the recommendations, and not just the health benefits.

Finally it is important that the guideline has been externally reviewed by experts prior to its publication and that there is a procedure provided for updating it in the future.

iv) Likely barriers and facilitators to implementation, strategies to improve uptake, and resource implications of the application of the guideline.

v) Editorial independence and whether the formulation of recommendations was biased with competing interests.

There are also potential adverse medicolegal consequences when the importance of methodology is overlooked. The proliferation of guidelines has happened at least in part due to a growing “accountability culture”.

The need to be able to refer to officially produced standards in any case where there is a legal challenge to medical practice has led to more and more guidelines to be produced.

Guideline methodology in such situations becomes even more important as a means to assess quality as there is the risk that guidelines of uncertain quality could be used medico-legally (both in and out of context) to dictate what a competent practitioner ‘would have done’ in particular circumstances.

Good quality guideline development aims to be evidence-based, systematic, and transparent. The purpose of guideline methodology is to ensure that this is the case. It provides the framework for standards in guideline development.

Guidelines are produced for many different diseases and encompass diverse diagnostics and interventions. Applying uniform methodology aims to provide standards of quality and transparency even though in some cases it creates methodological challenges for the developers.

It is imperative that anyone involved in healthcare is educated in the principles of evidence-based medicine and in particular in guideline development and guideline appraisal. This way each guideline can be assessed on its own methodology and development process in the setting within which it is going to be implemented.

It is always the responsibility of the individual clinician to decide on the individual patient's management. Guidelines aim to assist in this process but never to replace the medical acumen and experience of the specialist or the wishes of the patient. In order to assess the quality of a guideline the reader must be aware of the principles of evidence-based medicine and apply a systematic appraisal to the guideline.

Doctors should embrace and be educated to a high standard in the principles of evidence based medicine and especially in guideline appraisal so that they can responsibly assess for themselves the quality of any guideline and make a decision as to whether they will or will not apply it to their practice, or indeed change their practice to comply with guidelines.

Doctors must at all times be confident that they are doing the right thing for their patients and not adhering uncritically to guidelines.

The potential benefits of guidelines are only as good as the quality of the guidelines themselves. It is important to be able to assess the methodology used to develop the guidelines in order to be confident of the resulting recommendations.

5.0 Availability and quality of epidemiology data for a guideline needs assessment report

5.1 Introduction

A health needs assessment is a systematic method for reviewing the health issues facing a population. It aims to lead to agreed priorities and resource allocation that in turn aims to improve health and reduce any existing inequalities in health provision.[71] It is recommended practice for policy documents such as guidelines in order to inform and aid their better development as well as aid future strategic planning and implementation.[71]

With every NICE clinical guideline it is recommended that a baseline needs assessment be made. The scope of the colorectal cancer guideline required the first half of the guideline needs assessment to include information on the epidemiology of the disease.[69]

Epidemiological data is information on the factors affecting the disease in a way that makes it possible to infer possible trends. It is information regarding how the disease affects the population, the incidence, mortality, survival and prevalence.[72]

The use of epidemiological data on colorectal cancer for the purpose of a guideline needs assessment is to place the disease in the context of a cancer diagnosis in general, to make comparisons to other types of cancers, to present a focused picture of how the disease effects the population. This sets the scene for the guideline development group members called to shape the guidance.

Traditionally epidemiological information on cancer patients has been collected by the cancer registries. These are organisations for the systematic collection, storage analysis, interpretation and reporting of data on subjects with cancer.[73]

Population-based cancer registries collect data on all new cases of cancer occurring in a well defined population, usually resident in a particular geographical region. Their main objective is to produce statistics on the occurence of cancer in the defined population and to provide the framework for assessing and controlling the impact of cancer in the community.[73]

The data items collected vary depending on the resources available but generally include the basic information about the patient and the tumour and may extend to include information about the treatment and the follow up.[73]

The advancement of information technology has expanded the potential for data, storage, handling and analysis. The world wide web makes dissemination of any information much easier and faster than it has ever been in the past. As a result of both these changes the potential quantity of epidemiological information that could be accessible is large.

Though this is undoubtably a welcome improvement an important issue that must accompany the increasing quantity of information is that of the quality of the information. Two main issues need to be considered when evaluating the quality of epidemiological data: its completeness and its validity.[73]

Epidemiological data is collected in large databases and no such database is without error. According to international guidance regarding registration data all organisations handling such data should have quality control mechanisms to ensure continuous monitoring of data quality.[73,74]

5.2 Aim

To research the availability and quality of epidemiology data that can inform a guideline development group on incidence, mortality, survival and prevalence of CRC for the purpose of the CRC guideline needs assessment report.

5.3 Methods

Online data mining

Availability of Epidemiological data

The search engine 'Google' was used to search for available epidemiological data on cancer and CRC. The search terms used were: 'colorectal', 'cancer', 'incidence', 'mortality', 'survival', 'prevalence', 'epidemiology', 'UK'. The information produced from the search was selected based on whether:

i) the data related to the epidemiology of cancer and CRC in the UK.

ii) the data was supplied by a reputable source (e.g. national cancer registry, other national body with appropriate references to the source of the raw data used.)

iii) the data was collated, analysed, and presented in tables / graphs with accompanying explanatory text.

iv) the data was appropriate for presentation to non-medically trained individuals as the guideline development group also included lay persons.

The decision not to contact the registries directly for this information was taken *a priori* as cancer in general and CRC more specifically is a disease with a major population impact in the UK it was therefore anticipated that epidemiological data would be available in an analysed and collated format through the public domain / internet.

In addition a search for available data rather than a direct request for data would also provide the opportunity to scan the public domain for data of this

type potentially being provided by organisations other than the cancer registries.

Quality of Epidemiological data

The quality of epidemiological data was assessed by appraisal of the techniques used to create the data. Quality control is the mechanism by which the quality of the data is measured.[74]

There are national and international standards that all registries adhere to. The International Agency for Research on Cancer (IARC) and the International Association of Cancer Registries (IACR) have published and made publicly available documents that outline the parameters that affect data quality and the responsibilities of cancer registries.[73,74]

The UK Association of Cancer Registries (UKACR) plays a pivotal role in quality assurance of data provided by the registries through the development of national performance indicators. [75]

A systematic online search was carried out for information / reports regarding the quality control process as reported by the organisation providing the data. The search engine ‘google’ was used to search using the terms : ‘epidemiology’, ‘data’, ‘quality’ ‘control’ AND (a combination search using the term AND) the organisation identified in the availability search as providing the data. In addition the website of the relevant organisation was searched systematically by exploring every drop down list available through their home page and searching for epidemiology data quality reports.

If the organisation that presented epidemiology data identified in the availability search handled already-aggregated data and performed secondary analyses or presented the data for information purposes then the search was tailored to seek data regarding the quality of the analyses or the presentation of the epidemiological data (i.e. methodological information rather than quality control of the registration process).

5.4 Results

Availability of epidemiology data

Sources of epidemiology data

Cancer registries

In the UK epidemiological data on cancer is collected by the 8 cancer registries across England, as well as the national registries in Wales, Scotland and Northern Ireland. Amongst the cancer registries the Northern and Yorkshire Cancer Registry and Information Service (NYCRIS) has been appointed lead cancer registry for colorectal cancer in England. The Welsh Cancer Intelligence and Survival Unit (WCISU) provides colorectal cancer data for Wales. The Scottish Cancer Registry (SCR) and the Northern Ireland Cancer Registry (NICR) provide the colorectal cancer data for Scotland and Northern Ireland respectively.[75]
The registries produce incidence, mortality and survival summaries of their data for their geographical population catchment area annually and these are available as either text documents or summary tables from the website of each registry.

Office of National Statistics

In the UK cancer registries submit a standard dataset of information to the Office for National Statistics (ONS), for the collation of national cancer statistical data.[76,77] ONS is the UK's largest independent producer of official statistics and the recognised national statistical institute. [78] One of the key departments of ONS is involved in producing statistics covering life events (births, deaths and some health conditions). It is within this remit that cancer related statistics are produced. [79] These are published as reports,

bulletins or articles in the journal '*Health Statistics Quarterly*' all available through the publication section of the ONS website[80].

National Cancer Intelligence Network

The data collected by the registries is also accessed by other organisations, according to strict data access policies. The data is used for analysis and presentation / publication. One such organisation is the National Cancer Intelligence Network (NCIN). The NCIN was launched in 2008 and is a UK-wide initiative, working to drive improvements in standards of cancer care and clinical outcomes by improving and using the information collected about cancer patients for analysis, publication and research.[81] At the UK level, the NCIN co-ordinates information which is already aggregated by the registries. [82]

The NCIN brings data together into a National Cancer Data Repository (NCDR). The 1990 - 2010 England NCDR Analysis Dataset brings together data from each of the English Cancer Registries for the period from 1990 to 2010. The data consists of tumour level records submitted to the Office of National Statistics (ONS) by the English Cancer Registries together with a further sub-set of data covering additional data fields required for the purposes of analysis. The NCDR is held in a central location and is accessible by each of the cancer registries in line with NCIN's data access policies. Though NCIN includes English data only there is the capacity for the addition of the Celtic Countries NCDR Analysis Dataset. The creation of the NCDR2010 dataset is a joint project between the NCIN, UKACR and ONS.[83]

The NCIN uses the epidemiological data to coordinate UK-wide analyses as they become necessary. The NCIN annual reports include incidence, mortality, and survival figures and analyses that compare data across the UK by a variety of parameters (e.g. geographical location, age and others).[82]

It also uses the data to provide cancer information tools such as the cancer and prevalence e-atlases.[84,85] These aim to provide easily accessible basic information on incidence, mortality, survival [84] and prevalence [85] for the main types of cancers in males and females presented by UK region.

Cancer Research UK

Another organisation that handles the data from the registries and ONS is cancer research UK (CRUK), a leading UK charity. CRUK is dedicated to cancer research, and is funded entirely through public donations. Among their roles is providing publicly available relevant information necessary to understand the disease.[86]

The epidemiology data from the registries and ONS is available on the CRUK website presented for cancer in general and specifically for each cancer type including colorectal cancer. The information is presented in tables and graphs with accompanying text that is easily understood by non-medically qualified individuals.

CRUK has dedicated research teams including the Cancer Survival Group at the London School of Hygiene & Tropical Medicine (LSHTM) and the CRUK Statistics Team. They use the data to perform additional analyses as necessary, often in conjunction with the NCIN statistical teams.[86]

<u>Epidemiological data for cancer and colorectal cancer</u>

Life expectancy in the UK is increasing, due to a decrease in all-cause mortality. More elderly people are alive today than ever before.[87-92]

Cancer is a major cause of morbidity in the UK. The lifetime risk of cancer is an estimation of the risk that a newborn child has of being diagnosed with cancer at some point during his or her life. It is based on current incidence and mortality rates and therefore is calculated under the assumption that the

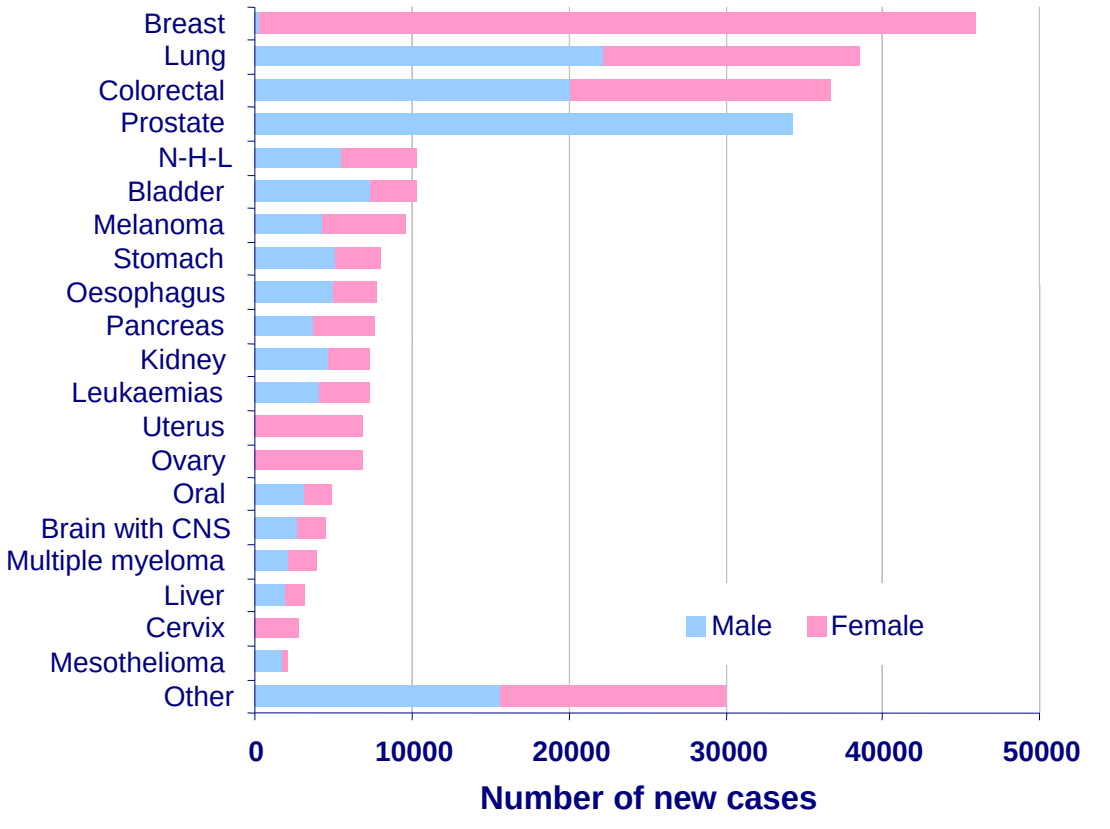
current rates (at all ages) will remain constant during the life of the newborn child. The lifetime risk of cancer in the UK based on 2006 data was 1 in 3. One in three people will develop some form of cancer during their lifetime.[87-93] It can develop at any age but is most common in older people. Around three-quarters of cases occur in people aged 60 and over (74%) and more than a third of cases in people aged 75 and over.[87-92]

Incidence of Cancer

Incidence of cancer refers to the number of new cancer cases arising in a specified period of time. Each year around 289,000 people are newly diagnosed with cancer. Breast, lung, colorectal, and prostate cancer account for over half of all the new cases (Figure 5.1).[94-97]

If current cancer incidence rates stay the same, by 2025 there will be 100,000 additional cases diagnosed per year due to the ageing population.[98]

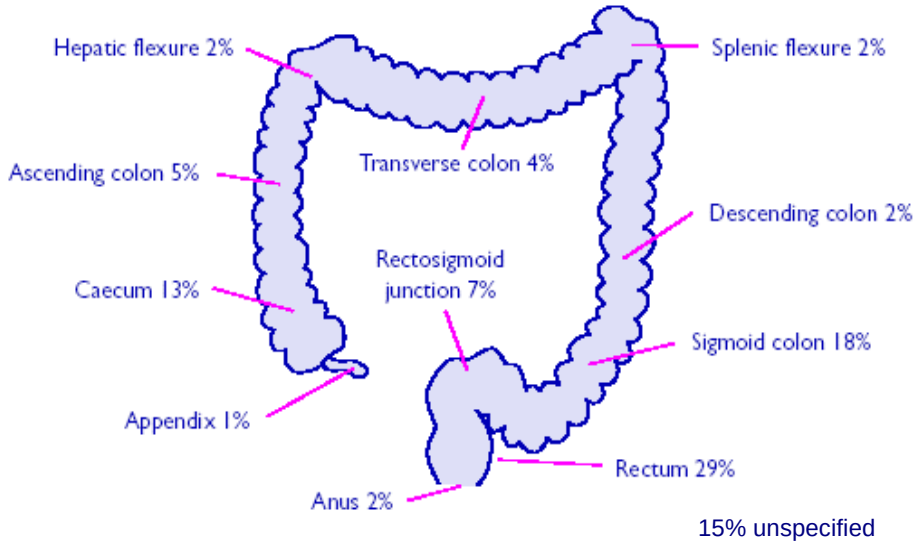
Figure 5.1: The 20 most commonly diagnosed cancers in the UK, 2005. (excluding non-melanoma skin cancer (nmsc)).[99]



Incidence of Colorectal Cancer

Colorectal cancer is the third most common cancer in the UK after breast and lung. Around 100 new cases of colorectal cancer are diagnosed each day. In 2005 there were 36,766 new cases of large bowel cancer registered in the UK - around two-thirds (22,748) in the colon and one-third (14,018) in the rectum. The left side of the bowel is affected by cancer more often than the right. Tumours in the sigmoid colon, rectosigmoid junction and in the rectum together account for over half of all cases (Figure 5.2).[100-103]

Figure 5.2: Percentage distribution of cases by site within the large bowel, England 1997-2000[104]



Colorectal cancer is the second most common cancer in women after breast cancer, with around 16,500 new cases diagnosed each year (Figure 5.3).[100-103]

More than 20,000 men are diagnosed with bowel cancer in the UK each year making it the third most common cancer in men after prostate and lung cancer (Figure 5.4).[100-103]

Figure 5.3: The ten most common cancers in females in the UK, 2005.[99]

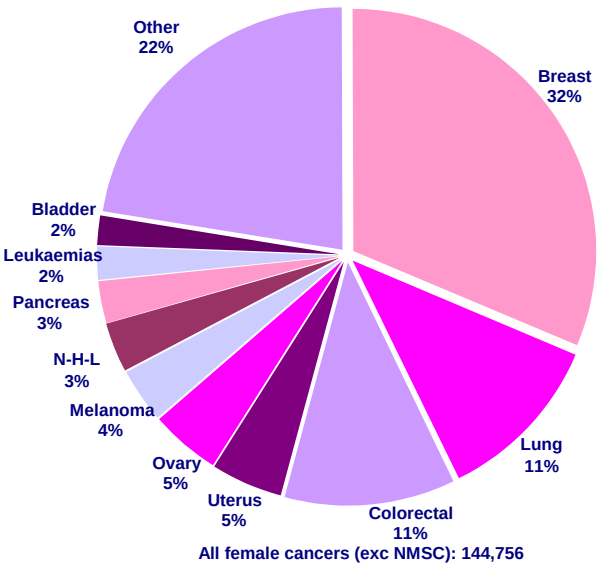
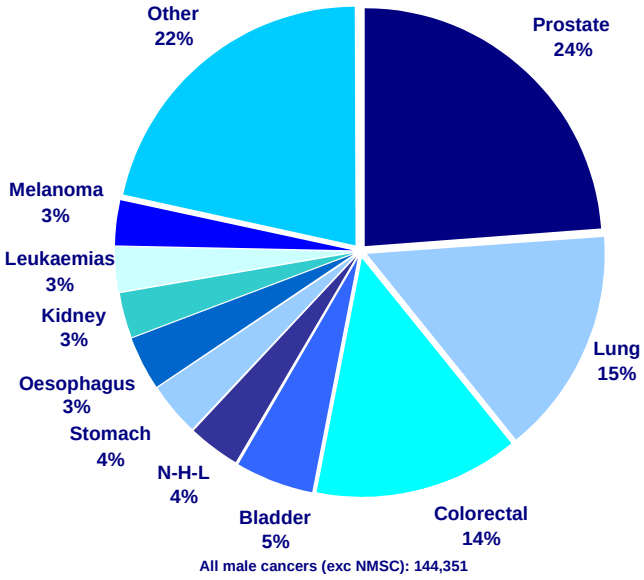
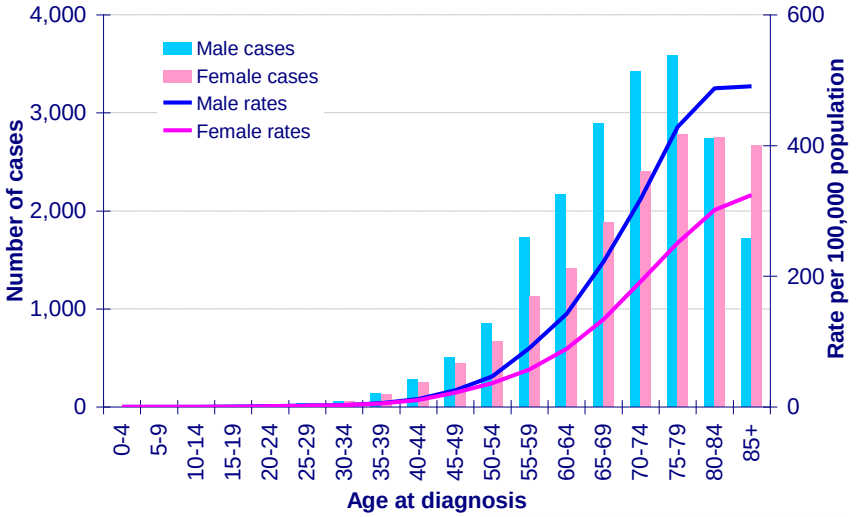


Figure 5.4: The 10 most common cancers in men in the UK, 2005.[99]



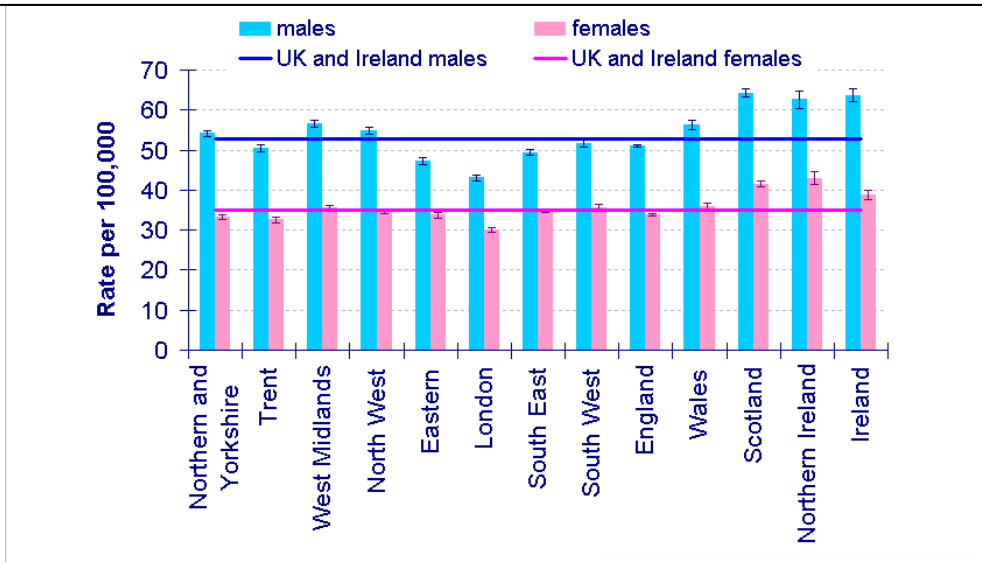
Almost three-quarters of bowel cancer cases occur in people aged 65 and over. Until age 50, men and women have similar rates for bowel cancer, but in later life male rates predominate. In numerical terms, there are more male cases of bowel cancer up to the age of 80, after which female cases are in the majority, even though their rates are lower, as women make up a larger proportion of the elderly population (Figure 5.5).[100-103]

Figure 5.5: Numbers of new cases and age-specific incidence rates by sex, bowel cancer, UK, 2005.[105]



Geographical analysis of cancer incidence in the UK showed similar distribution for colon and rectum with small variation (Figure 5.6).[100-103]

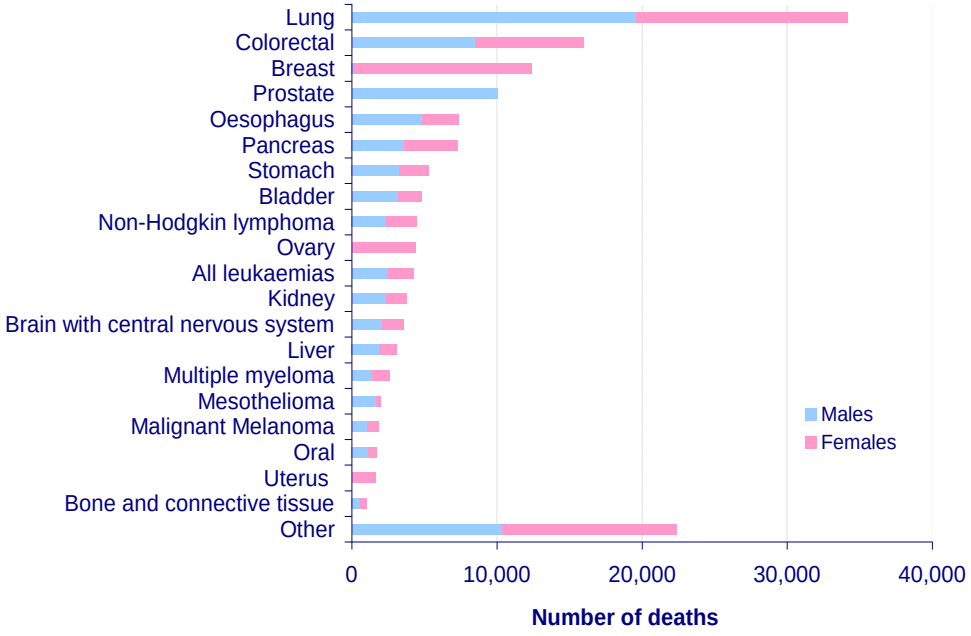
Figure 5.6: Age-standardised incidence rates by sex, colorectal cancer, region of England, UK and Ireland, 1991-1999 [105]



Mortality from cancer

In the UK in 2006, there were 154,162 deaths from cancer: one in four (27%) of all deaths in the UK; 29% for males and 25% for females. Deaths from cancers of the lung, bowel, breast and prostate together account for 47% of all cancer deaths (Figure 5.7).[106-108]

Figure 5.7: 20 most common causes of death from cancer, UK, 2006.[109]



Cancer caused a quarter of deaths in the over 65s in the UK in 2006, whereas cancer was responsible for more than a third (36%) of all deaths in the under 65s.[106-108]

In females under the age of 65 cancer causes 45% of deaths, while in males it is only 30%.[106-108]

The overall cancer death rate has fallen by 10% over the last decade around 12% for men and 9% for women.[106-108]

The majority of deaths from cancer occur in the elderly. More than three quarters of cancer deaths (76%) occur in people aged 65 years and over. [106-108]

The cancer death rates rise with increasing age. Although there is a higher number of cancer deaths in the over 65s, cancer causes a greater proportion of deaths in younger people. [106-108]

Mortality from Colorectal Cancer

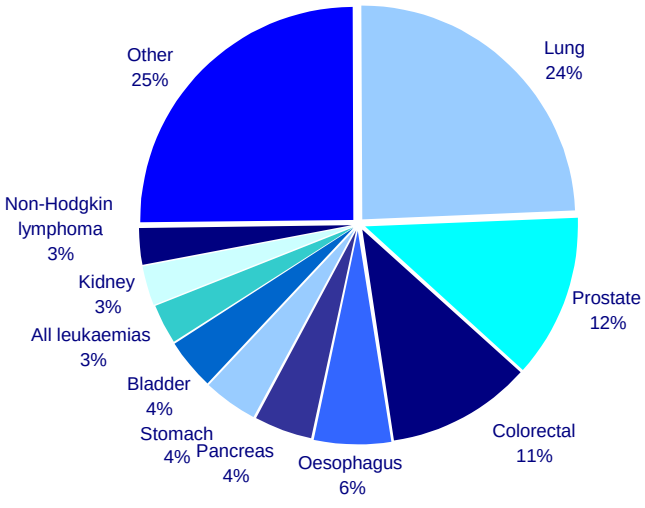
Colorectal cancer was the second most common cause of cancer death (10%) after lung cancer (Figure 5.7).[110-112]

In 2006 there were 15,957 deaths from colorectal cancer in the UK, comprising 10,119 from colon and 5,838 from rectal cancer.[110-112]

Colorectal cancer caused 8,511 deaths in men in 2006, accounting for 11% of all male cancer mortality (Figure 5.8).

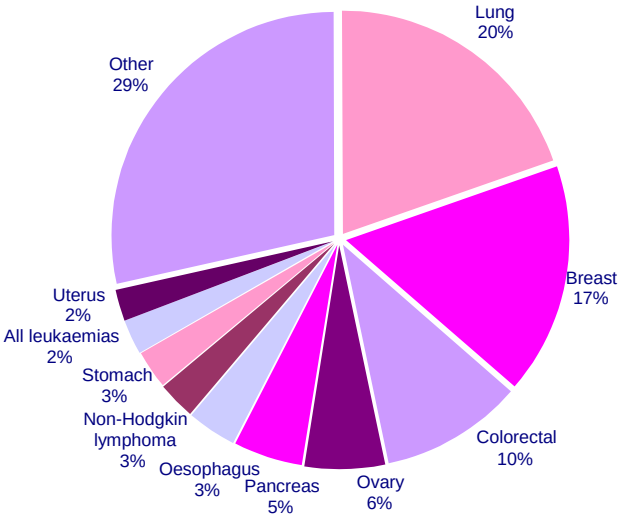
Colorectal cancer was responsible for 7,446 deaths and 10% of all cancer deaths in females (Figure 5.9).[111-113]

Figure 5.8: The 10 most common causes of cancer deaths, males, UK, 2006.[110]



All malignant neoplasms

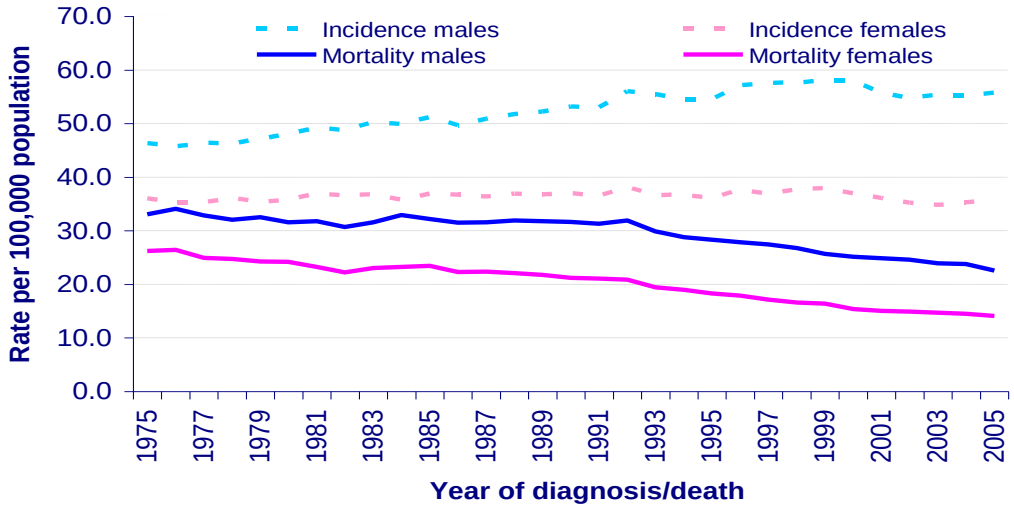
Figure 5.9: The 10 most common causes of cancer deaths, females, UK, 2006.[110]



All malignant neoplasms

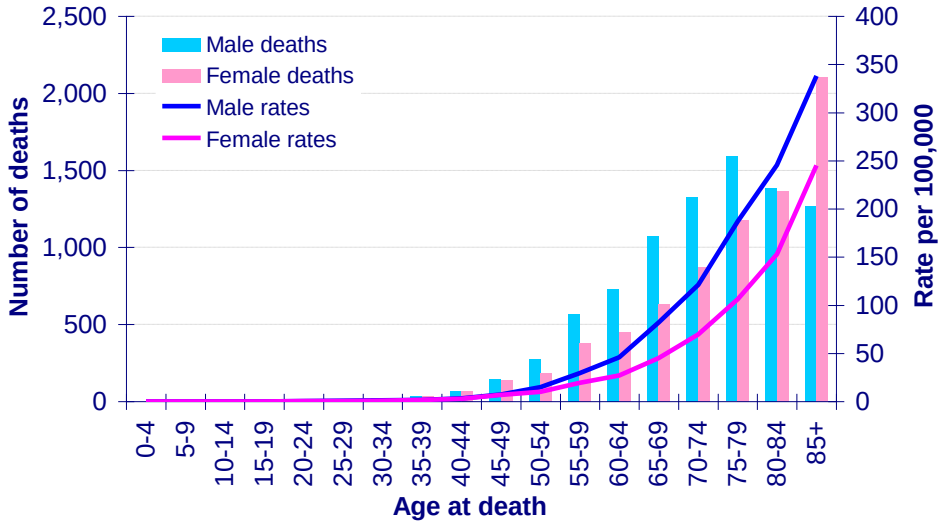
In contrast to incidence trends, bowel cancer mortality has been falling fairly continuously since the early 1990s (Figure 5.10).[110-112]

Figure 5.10: Age-standardised incidence and mortality rates by sex, colorectal cancer, Great Britain, 1975-2005.[113]



80% of bowel cancer deaths occurred in people aged 65 and over and almost two-fifths in the over 80s (Figure 5.11). [110-112]

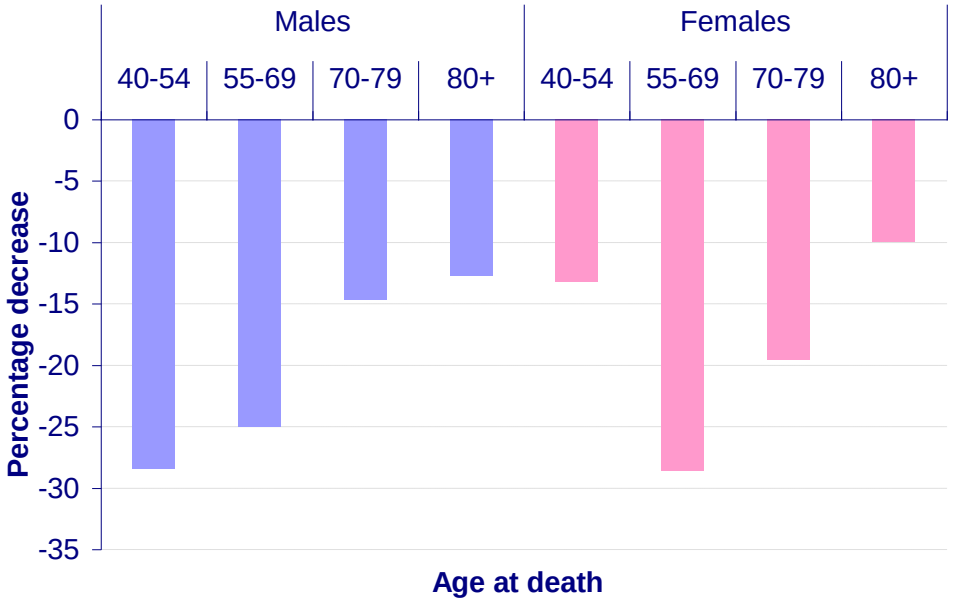
Figure 5.11: Number of deaths, and age-specific mortality rates, colorectal cancer, by sex, UK, 2006.[113]



In the ten years between 1997 and 2006, the bowel cancer age-standardised mortality rates in the UK fell by 17%. This fall in mortality affected all age groups with the largest fall in the 40–69 age groups for men and the 55-79 age groups for women. [110-112]

Bowel cancer mortality rates started to decrease in 1988 and since then the male rate has fallen by 30% and the female rate by more than a third (36%) (Figure 5.12).[110-112]

Figure 5.12: Percentage decrease in mortality rates, bowel cancer, by age and sex, UK, 1997-2006.[113]



Within England, bowel cancer mortality rates are generally higher in the north of the country. [114]

Survival from Cancer / Colorectal Cancer

Survival estimates are the percentage of patients who are still alive a specified time after their diagnosis of cancer. The most common estimates are five and ten year survival rates.[115]

Relative survival provides an estimate of the percentage of patients still alive a specified time from their diagnosis, taking into account the background mortality in the general population (i.e. the percentage of patients that would be expected to have died from other causes during the period if they did not have cancer). It is therefore an estimate of the proportion of patients that survive their cancer for the specified time period.[115]

Survival has improved for most cancers in both sexes during the 1990s.[116-119] There have been similar and significant improvements in survival for both colon and rectal cancer over the last 25 years.[120] The five-year relative survival rates for both male and female colon and rectal cancer have doubled between the early 1970s and early 2000 (Figures 5.13, 5.14, 5.15, 5.16).[121-123]

Five-year relative survival for male colon cancer rose from 22% in the early 1970s to 52% in early 2000; for females it rose from 23% to 53%. Five-year survival rates for male rectal cancer rose from 25% in the early 1970s to 50% in early 2000 and from 27% to 52% for female rectal cancer. On average, increases in five-year survival of around 4% every five years for colon cancer and around 5-6% for cancer of the rectum occurred in both men and women.[121-123]

Ten-year survival rates are only a little lower than those at five-years indicating that most patients who survive for five years are cured from this disease.[123] These improvements are a result of earlier diagnosis and better treatment but there is still much scope for further progress.[121-123]

Figure 5.13: Age-standardised relative survival in men diagnosed with colon cancer, England and Wales, 1986-1999.[124]

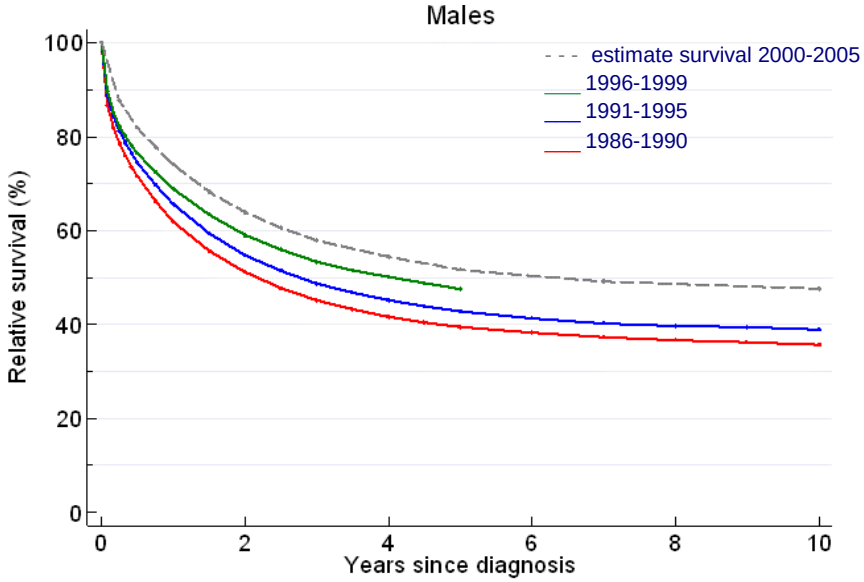


Figure 5.14: Age-standardised relative survival in women diagnosed with colon cancer, England and Wales, 1986-1999.[124]

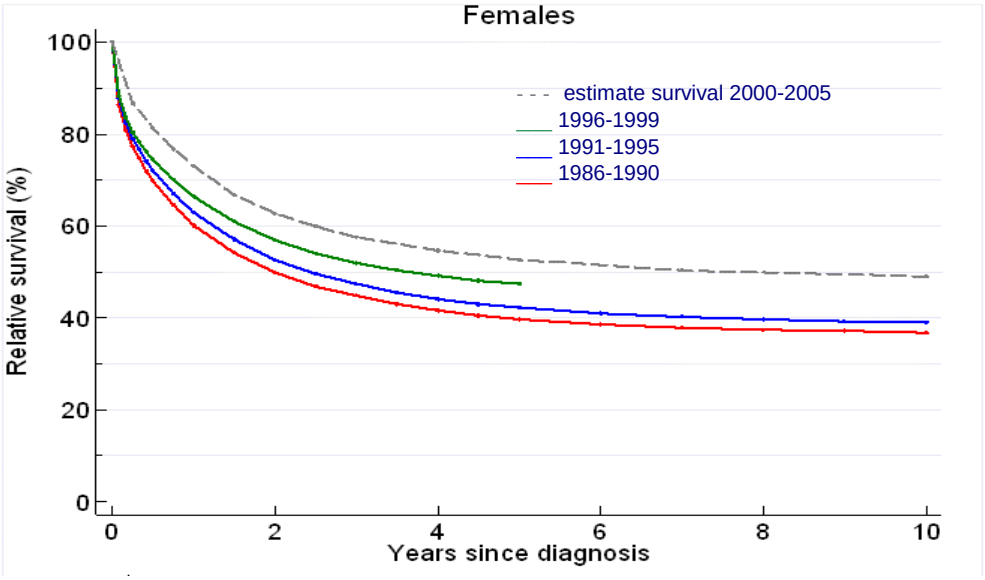


Figure 5.15: Age-standardised relative survival in men diagnosed with rectal cancer, England and Wales, 1986-1999.[124]

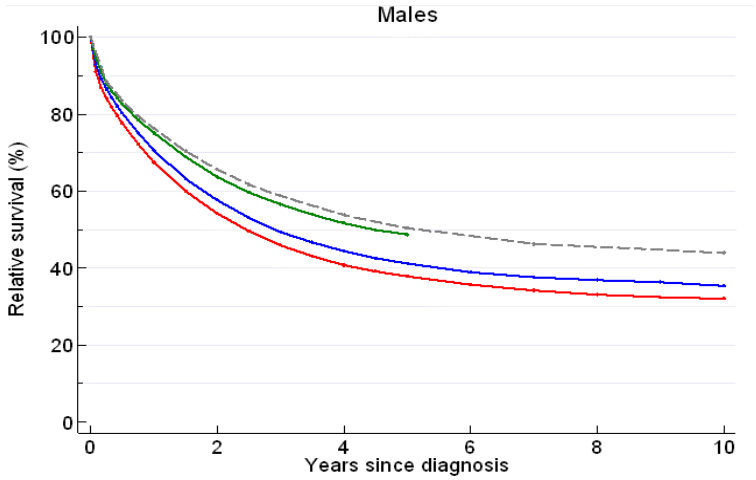
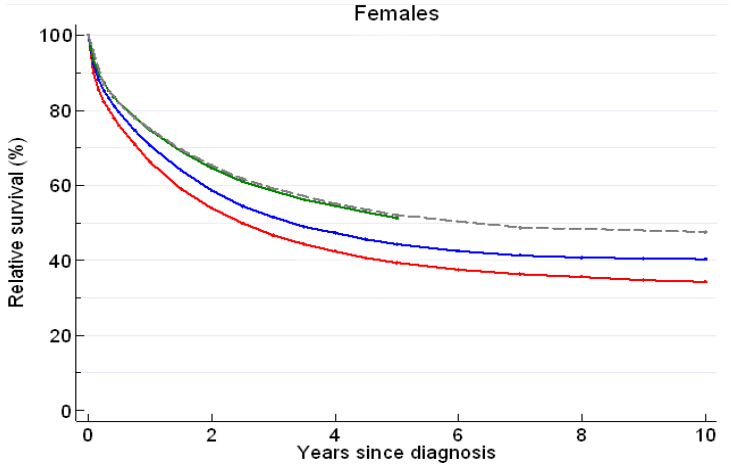


Figure 5.16: Age-standardised relative survival in women diagnosed with rectal cancer, England and Wales, 1986-1999.[124]

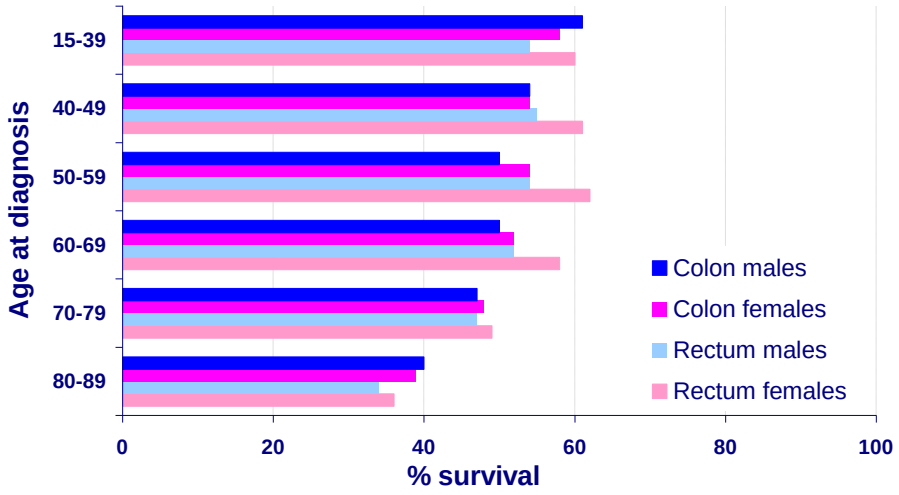


Younger bowel cancer patients have a better prognosis than older patients.[121-123] As with nearly all cancers, relative survival for bowel cancer is higher in men and women under the age of 70, even after taking account of the higher background mortality in older people (Figure 5.13).[121-123]

The reasons for this are likely to include a combination of better general health, more effective response to treatment and earlier diagnosis in younger people overall. Differences in underlying tumour biology may also play a part for some cancer sites.[125]

Five-year survival in the age group 60-69 is slightly higher than the 40-49 and 50-59 age groups though this difference is not statistically significant (Figure 5.17).[123-125]

Figure 5.17: Five year survival (%) of patients diagnosed with bowel cancer 1996-1999, England and Wales.[125]



Patients who are diagnosed at an early stage have a much better prognosis than those who present with more extensive disease. Over 83% of patients diagnosed with Dukes stage A survive five years compared with less than 3% of patients with advanced disease (Table 5.1).[126]

Table 5.1: Approximate frequency and 5 year relative survival (%) by Dukes stage – England and Wales,1996-1999[127]

Dukes stage	Approximate frequency at diagnosis	Approximate 5-year survival
A	11%	83%
B	35%	64%
C	26%	38%
Metastatic disease	29%	3%

Prevalence for cancer and colorectal cancer

Cancer prevalence refers to the number of people who have previously received a diagnosis of cancer and who are still alive at a given time point. Some of these patients will have been cured and others will not. Therefore prevalence reflects both the incidence of cancer and its associated survival pattern.[128]

Overall, it is estimated that there are now 2 million cancer survivors in the UK, or approximately 3.3% of the population of the UK (Table 5.2).[128] This figure is rising at an estimated 3.2% per year. Overall, 10% of the total UK population over the age of 65 years is now a cancer survivor.[128]

Table 5.2: UK estimates of total cancer prevalence. (UK 2008 estimates based on diagnoses 1971-2004 applied to 2008 population; Thames cancer registry 2008.)[130]

Breast (female)	550,000
Large bowel	250,000
Prostate	215,000
Lung	65,000
Other	920,000
All cancers	2,000,000

These latest estimates are much higher than previous forecasts of cancer prevalence.[129] This is mainly because incidence has been rising whilst the death rates have continued to fall, leading to better survival. This trend is expected to continue over the coming years as a result of a number of factors, including an ageing population, earlier detection of cancer and continued improvements in treatment.[128-129]

As the incidence of bowel cancer is high and survival rates have doubled over the last 30 years there are many people alive today who have been diagnosed with bowel cancer. An estimated 250,000 people are alive in the UK having received a diagnosis of bowel cancer (Table 5.2).[128]

The NHS Bowel Screening Programme which has now been implemented nationally will dramatically influence the epidemiology of the disease and it will increase prevalence with more patients being diagnosed earlier and at an earlier stage giving them better prognosis and therefore increasing the prevalence of the disease. There could be up to 20,000 fewer deaths from bowel cancer over the next 20 years if just 60% of those eligible take up the invitation for bowel screening.[128]

Quality of epidemiological data

UK registry data quality

Each Cancer Registry in the UK provides quality control information annually to the UKACR on a number of measures to allow comparisons of the timeliness, quality and completeness of their data. This information is collated and an annual report produced by the UKACR, consisting of a series of datasets, with accompanying explanatory commentary from the registries.[76] The UKACR annual quality results for 2008 from the 4 main registries providing colorectal data are presented in Table 5.3.

Table 5.3: UKACR quality and performance indicators 2008 and results from the 4 registries that provide colorectal cancer data

Quality and performance indicator	Cancer peer review standard	Target expected	UK average	NYCRIS England	WICSU Wales	SCR Scotland	NICR N Ireland
1. Registration and timeliness	100% +/- 2%	100% +/- 2%	95.7%	95.7%	103.1%	48.2%	103.3%
2A. % change in registrations – male colorectal all xnmsc	+/- 2%	+/- 2%	 1.1 1.2	 0.8 -3.2	 5.9 3.4	 -1.3 -2.2	 2.1 5.7
2B. % change in registrations – female colorectal all xnmsc	+/- 2%	+/- 2%	 1.4 0.6	 -1.2 -4.7	 6.9 2.9	 -0.8 -0.5	 6.6 0.9
2C. childhood cancer incidence rates	na	na	na	na	na	na	na
2D. % death certificate only cases (DCO) colorectal	 2%	 2%	 2.3%	 0.4%	 1.9%	 0.6%	 0.4%

2E. % “True” DCO colorectal	2%	2%	1.4%	1.0	1.9	0.6	0.4
2F. % zero survival colorectal			3.3%	2.2%	2.5%	1.3%	0.7%
2G. % microscopically verified colorectal	87%	88%	88.9%	90.8%	84.9%	91.4%	90.5%
2H. % non-specificity of morphology code for cases microscopically verified colorectal			2.5%	1.6%	5.7%	1.1%	3.5%
2I. mortality:incidence ratios colorectal		0.44	0.43	0.44	0.43	0.45	0.39
3A. demographic and diagnostic details patients name patients address sex ethnicity date of death postcode	 100%		 100% 99.9% 100% 50.2% 100% 99.8%	 100% 100% 100% 42.3% 100% 100%	 100% 100% 100% 30.2% 100% 100%	 100% 100% 100% 0.2% 99.9% 100%	 100% 99.1% 100% 0.0 100% 97.8%

date of birth	100%		100%	100%	100%	100%	100%
unique health identifier	100%		96.7%	99.9%	99.9%	91.3%	81.1%
diagnosis date	100%		99.7%	98.7%	100%	100%	100%
site of primary	>95%		96.2%	95.3%	96.5%	96.1	95.9%
type of growth	>85%		87.4%	87.3%	84.4%	87.8%	84.6%
behaviour of growth	>98%		99.9%	99.4%	100%	100%	100%
basis of diagnosis	>96%		97.7%	99.3%	89.7	99.3%	92.7%
3B/C. treatment/ screening/stage information							
therapeutic surgery			45.6%	43.6%	54%	43%	46.3%
radiotherapy			17.8%	20.5%	na	19.9%	15.8%
chemotherapy			19.5%	24.0%	20.2%	26.1%	20.6%
Dukes stage	>74%		66.4%	75.9%	52.6%	76.4%	55.4%

When the search for quality control data was performed directly via the registry websites the only registry that provided the results of their quality control in tabulated format on their website was the NYCRIS registry.

NYCRIS data quality

NYCRIS the lead registry for colorectal cancer registrations in England produces an annual report which includes reporting of data quality. The report is presented and is easily accessible through the NYCRIS home webpage. The website also contains information regarding the quality assurance process in place at NYCRIS and this information is summarised below.

NYCRIS has in-house quality assurance processes that aim to maintain high quality data. These comprise extensive routine monthly quality checks as well as one-off excercises and spot checks. In addition there is a comprehensive quality assurance process for reviewing each successive year of completed registrations.[131]

Quality control procedures of the registration process include routine checking of staging against staging protocols, of registrations with multiple hospital episodes, and of registrations where two or more treatments are recorded.[131]

NYCRIS uses monthly monitoring reports to estimate the completeness of a particular registration year based on the numbers of registrations made in previous years. NYCRIS liaises with other organisations in the geographical area that shows under-ascertainment (a commonly occuring problem) and identifies alternative means of retrospective cross-checking of cases, ideally prior to the completion of a particular registration year. This often requires pathology reports to check for missed cases. It is not clear why some cases are missed in routine process; changes in hospital personnel or changes in computer systems may have an impact. Reminders are sent regularly regarding the range of registerable conditions and monitoring of this aspect of data quality continues.[131]

Table 5.4 presents the key quality and performance indicators for NYCRIS compared to the average UK results based on cases diagnosed in 2007 as reported in the NYCRIS 2008-9 report available through the NYCRIS website.[132]

Table 5.4: NYCRIS performance indicators for 2007 diagnoses.[132]

Performance indicators	NYCRIS	UK average
Dataset completeness	Data item % complete	Data item % complete
Postcode	100.0	99.9
Sex	99.9	100.0
Date of Birth	100.0	100.0
NHS number	99.4	99.0
Topography (specific)	95.9	96.4
Morphology (specific)	86.7	88.4
Breast staging	Na	49.0
Cervix staging	92.1	64.7
Colorectal staging	75.8	68.8
Melanoma staging	92.9	58.4
Grade (breast only)	89.9	86.3
Death Certificate only cases	%	%
All sites (male)	1.9	2.3
All sites (female)	1.9	2.4
Microscopic verification rates	%	%
All sites (male)	86.1	84.4
All sites (female)	90.2	85.9
Mortality : Incidence ratios		
All sites (male)	0.55	0.53
All sites (female)	0.53	0.50

ONS data quality

ONS provides an extensive list of methodology and guidance documents regarding the statistical analyses it carries out. These are available to access through its webpage.[133,134]

NCIN data quality

The NCIN uses the data already aggregated by the cancer registries and ONS. For any additional statistical analyses performed the methodology is provided as part of the introduction to the analysis document. All analyses are available through the website.[135]

CRUK data quality

CRUK uses the data provided by the UK cancer registries and ONS. It does not provide any information on the quality of the registration data or the analyses from ONS. However it provides explanations on terminology used as part of the explanatory text accompanying the statistics.[136]

5.5 Discussion

The results show that epidemiology data on colorectal cancer is available and easily accessible through multiple sources via the world wide web. These sources are long established national organisations with an expertise in the data being handled. Therefore they are appropriate sources for delivering this information. In addition, the information is available without prior request. Either in graphical format or explanatory text information on the epidemiology of cancer and colorectal cancer can be accessed easily by both professionals and the public. This fact in itself makes the issue of the quality of the data and the ability to access and assess the quality of the data even more important.

Quality is a property of the data and a product of the techniques used to create the data. [74] The main source of data is the cancer registries that perform the main task of cancer data registration. Therefore the assessment of quality begins by the assessment of the quality of the registration process.

The main sources of data for a cancer registry include: 1) treatment facilities such as hospitals, hopsices, GP surgeries; 2) diagnostic services, especially pathology departments, imaging departments, haematological laboratories; 3) death certificates from the death registration system.[73]

The information is collected both actively (registry staff visit the hospitals to collect information) and passively (hospital staff fill in registration forms provided by the registry for each cancer case). A mixture of both procedures, with an emphasis on the latter is followed in most registries creating large databases of information.[73]

There can be error in any of the steps of the information gathering and processing exercise. No large-scale database can be perfect and quality control procedures are instituted to identify the areas and degree of imperfection, and thus assist in the interpretation of the data and any indicated procedural changes.[74]

The results show that in the UK all registries have quality control procedure in place and aim to adhere to standards set by national and international bodies. This in itself is an important indication of quality. There is monitoring of the process irrespective of the results of the quality control process itself.

With regard to the quality of the data two main issues need to be considered when evaluating the quality of the data of a cancer registry: its completeness and its validity.

Completeness of the data includes completeness of cover. A population-based registry should, by definition, register every single case that occurs in its catchment population.[73] It is also important to ascertain the extent to which the registry eliminates registrations of cases from outside the catchment population and avoids multiple registrations of the same person or tumour.[73] Data completeness also included the completeness of detail i.e. ascertaining every item of data for every patient. Some data items may not be

applicable to every patient, there may be errors of commission (data being present where it should be absent), or errors of ommision, all making interpretation of the data difficult.[74]

Validity of the data refers to the accuracy of the detail recorded. Error can occur in a multitude of ways: abstraction, transcription, coding. In additon, validity of the data refers to the accuracy in the reporting. Many variables, discontinuity of coding, changing file layouts, make the collation of data a difficult task. All this may be handled by staff that lack first-hand knowledge of the data and this may introduce error that is difficult to detect unless it gives rise to totally unexpected results. Finally, validity of the data also refers to the accuracy of the interpretation of the data. This requires understanding of the data, the data sources and how the data is processed.[74]

The results show that UK cancer registries have developed internal quality control checks so that attention is drawn to missing information and inconsistent data.

Case completion and ascertainment is rarely complete. Various methods, such as comparisons with death certificates and hospital records have been used to determine the degree of completeness of registration.[73]

A unique registration number is assigned by the registry to each patient. If a patient has more than one primary tumour then the same number is given to each tumour. Multiple primaries are then distinguished on the basis of their incidence date and their topography (site of primary tumour) and morphology (histological type of the tumour). The incidence date is the date of first consultation or admission to hospital with a diagnosis of cancer as can be verified from the hospital records. If this is not available then it is the date of the first pathological report that confirms cancer. A special problem arises if cancer is first ascertained from a death certificate and attempts to follow back are unsuccessful. The date of death of such ‘death certificate only’ (DCO) cases is taken as the incidence date.[73]

The results show that all four registries had a very low death certificate only case load and less than the target of 2%.

Other identification items such as name, sex, and date of birth are important to avoid multiple registrations of the same patient or tumour, to obtain follow up data and to conduct any kind of linkage. Address is essential to conduct analyses by area of residence.[73] The results show that all four registries had a very high completion rate for these parameters between 99-100%.

An important deficiency is the recording of ethnicity. This parameter is important where distinct ethnic groups might carry different risk of cancer. Ethnicity registration varied between 0.2 and 52% which is a much lower registration completion than any of the other parameters. It affects all registries, which indicates that this is a national problem regarding the registration of this particular parameter rather than poor quality in any one step of the registration process of one registry. Further analyses of colorectal cancer epidemiology in combination with ethnicity data would need to take this into account but it does not effect the data within this analysis.

The validity of the data can be assessed in various ways. The proportion of cases with microscopic verification of diagnosis is a very useful index, as is the proportion registered during life (not simply from death certificate).[73]

The results show that all four registries have high proportion of cases with microscopic verification (between 84.9% and 91.4%). The Welsh registry with a score of 84.9% is slightly under the standard target of 87% and under the UK average of 88.9%; however overall these are high scores.

Information on the most valid basis of diagnosis is of great interest in assessing the quality of the registration data. The minimum requirement of a cancer registry is to discriminate between tumours that were microscopically verified and those that were not. If possible, further information should be obtained to distinguish neoplasms that were diagnosed on the basis of a

clinical history only, clinical history plus other investigations (e.g. x-ray), exploratory surgery, autopsy, cytology, etc.[73].

The results show that all four registries had high percentage of registration of this parameter (89.7-99.3%). The Welsh registry was scoring below the standard target of 96% and less than the UK average of 97.7. However this is a steady improvement from previous years (84.6% in 2006, 88.5% in 2007).

NYCRIS, was the only registry that actually displayed its quality control results on its webpage. This is good practice and certainly makes assessment of data both easy and possible concurrently when approaching a source for data.

Overall, the quality of UK cancer and colorectal cancer epidemiology data is high primarily due to the high quality of the cancer registration process in the UK carried out by the cancer registries and the quality control procedures in place.

The establishment of the English National Cancer Online Registration Environment (ENCORE) as a single database recording all English cancer registrations will reduce the potential for error arising from the collation of information from multiple databases across the various English registries as well as provide a single point of contact for accessing the epidemiological data of interest.

The quality of the data is also a reflection of the quality of the statistical analysis carried out by ONS and to a lesser degree by the statistical teams at CRUK and NCIN. The results show that these organisations present their methodology clearly through their websites. Either as separate documents (in the case of ONS), or as part of a specific analysis document (in the case of NCIN), or as part of their ‘frequently asked question’ webpage in the case of CRUK. The transparency of the methods of statistical analysis is very important.

It is important to know why a particular method has been used over an alternative and how a particular calculation has been carried out.

The results present the data as age-standardised rates for incidence and mortality and it is on these age-standardised rates that survival and prevalence are estimated thereafter.

Crude incidence rates are calculated using a simple formula in which the number of cases is divided by the corresponding population and multiplied by 100,000. Since cancer is generally more common in the elderly, crude rates are greatly influenced by the proportions of older people in the populations being studies. For this reason, age-standardised rates are used when making comparisons of incidence rates (for example, over time, between sexes or between geographical areas).[137]

Age-standardisation adjusts rates to take into account how many old or young people are in the population being studied. Thus when rates are age-standardised, differences in the rates over time or between geographical areas do not simply reflect variations in the age-structure of the population. This is important when looking at cancer rates because cancer is a disease that predominantly affects the elderly. So if cancer rates are not age-standardised, a higher rate in one country is likely to reflect the fact that it has a greater proportion of older people.[138]

The data and the analyses presented do have limitations. All the epidemiology data presented has a time lag of about two years from the time of the search query. This is a common and accepted reality in data of this type. The process of registering a cancer is complex and there are a number of processes in place as discussed above to ensure the data is of high quality but this means there is usually a delay of about 18 months before the data is complete enough for them to be published. In addition, the statistical analyses are compiled from data produced by the regional cancer registries in England and the three national registries in Wales, Scotland and Northern Ireland which means there is another time lag before the statistical analyses can be

published as these have to wait until all of the data has been published by each country and its registry.

In addition, mortality may be higher than case numbers because of the way death certification and cancer registration works, and what data is available and when. If a patient has died from cancer but the official documenting the death on the certificate cannot confirm what type of cancer caused the death, it may be recorded as a non-specific cancer type (Cancer of Unknown Primary -CUP). However, on receipt of the death certificate the cancer registries may then be able to identify other information about that history and determine what type of cancer it was, and update the record regarding their case diagnosis. The data would therefore show a patient recorded as having a known type of cancer for their case data, e.g. being a colorectal cancer case in the incidence data, but as a different cancer type in the mortality data, e.g. being a CUP death in the mortality data. This inconsistency will remain because the death certificate cannot be changed and the effect of this is potentially inflated mortality statistics for non-specific cancers like CUP at the detriment of the mortality statistics for specific cancers.[138]

The epidemiology results were presented by the author to the guideline development group as part of the guideline needs assessment report.
The results trigerred a discussion about the implication of these results internationally and how the UK compared to other countries. In the past studies have shown that UK nations have poorer cancer survival outcomes than comparable countries. However, the opinion of the GDG was that such international comparisons lacked uniformity in the countries they compared and this information should therefore not be searched nor included in the guideline document.

Around the same time the International Cancer Benchmarking Partnership (ICBP) was being set up (2009) and its purpose mirrors the conclusion of the GDG on this matter. The ICBP is an international initiative involving 6 countries. Australia, Canada, Denmark, Norway, Sweden and the UK. The partners all have comparable wealth, a universal access to health care and long standing high

quality population based cancer registration. Partner countries are invited to join only if they fulfil these three criteria. Meeting these three factors is important as it limits the variables that can affect cancer outcomes. The research is looking specifically at four cancers breast, colorectal, lung and ovarian which share a large burden of cancer disease in developed countries and which also display significant differences in survival amongst countries.[258] Comparing cancer services, processes and the public's interactions with these across different countries can help identify possible reasons for the observed differences. And by understanding critical differences and similarities between UK nations and those countries with better survival rates it should be possible to build a picture of where improvements could be made.[258]

Overall what the results show is that a reliable source of epidemiological data is a powerful tool that can provide clues as to how the lives of people with cancer are affected and can be changed. This can be used to improve early detection, diagnosis, treatment and follow up. As part of a guideline document it can either guide the direction of research or underpin future improvements based on specific findings.

6.0 Availability and quality of current clinical practice data on CRC management for the purpose of a guideline needs assessment report

6.1 Introduction

The scope of the NICE colorectal cancer guideline recommends that the second half of the needs assessment report includes information on current UK clinical practice with regard to the management of colorectal cancer.[139]

The aim is to identify any concerning variability that exists in the management of the disease in order to help the guideline development group members formulate recommendations that are likely to have the greatest impact on clinical outcomes. [139]

Variability in disease management is not always undesirable. It is often just a reflection of different populations, patient choice, or one center of excellence carrying out a technique in a slightly different way to another center of excellence.[71]

It is unacceptable however if patients receive substandard care which deviates from basic principles due to the actions and practice of one individual clinician without backup and support from a multidisciplinary team.[71]

In the UK variability in disease management is likely to exist in areas such as pre-operative staging, often because of the different availability of imaging resources.[71]

Also, where a sequence of therapeutic interventions is possible the sequence may differ, e.g. treatment for liver metastases either before or after the surgery for the primary tumour. [71]

Variation is also likely to exist when considering different patient groups. For example when considering patients with multiple co-morbidities or the very elderly

who despite their age have low morbidity as assessed on the ASA physical status classification system. Some do and some do not get access to radical therapies.[71]

The collection of clinical practice data is a relatively recent development. Before 1987, only a 10% sample of admitted patient records were collected nationally. [140,141]

The mechanisms for collecting the data have changed considerably over the years, often in response to changes in the organisation of the NHS.[140]

There are additional clinical practice data sources that have been developed in recent years, such as national audit databases, disease registers, specialty datasets, and national specialist systems, all of which have been developed to provide high-quality clinical data and improve the quality of care provided.[141]

When considering the quality of current clinical practice data the two main issues of concern are the completion and the validity of the data. Both these parameters are intimately involved with the process of the data collection. Understanding the data collection process, its strengths and weaknesses, and having quality control processes in place for monitoring are vital in order to ensure data that is reliable.

The issue of the quality of current clinical practice data is complicated however by the sheer diversity of this type of data. The data can range from imaging information, to operative information, to patient reported outcomes and many more types.

Furthermore, quality issues are additionally complicated by the diversity of the databases. The location of these, their age, the software, the data collection process they encompass, the data access policies are some of the many variables relating to the databases. In order to ensure quality assurance a systematic approach with attention to the elimination of bias risk is key.

6.2 Aim

To research the availability and quality of current clinical practice data that can inform a GDG on aspects of CRC management for the purpose of the CRC guideline needs assessment report.

6.3 Methods

For the purpose of a needs assessment current clinical practice data could potentially be sourced from many different specialist areas within the different settings of colorectal cancer management. In order to limit the breadth of this analysis the scope of the guideline was used as a guide to defining what the required current practice data should include.

As the purpose of the needs assessment is to 'set the scene' for what is currently happening in the UK in colorectal cancer management it was deemed to be useful to identify sources of current clinical practice data relevant to each clinical topic the guideline was planning to address.

6.3.1 Formulation of data queries (current clinical practice questions):

The topics for the clinical questions that the guideline addresses are based on the guideline scope and are developed in consultation with the various stakeholders at the scoping workshop.

Once the specific guideline topics emerged after the results of the scoping workshop, the guideline development group together with the NCC-C technical team (including the author) finalised the fifteen clinical questions for the NICE colorectal cancer clinical guideline.

A background summary of key facts known about each clinical question was presented to the guideline development group by the lead group member for this

topic as chosen by the guideline chair. This was for the purpose of establishing the basic information available about each one topic prior to any systematic review performed.

Each guideline question together with the background information provided was then used by the author as the basis from which the equivalent current clinical practice data query was formulated.

6.3.2 Online data mining

With the current practice questions formulated an online search was carried out to identify appropriate sources that could provide data to answer these specific clinical practice questions.

The search was through the ‘Google’ search engine and used key words from the formulated current clinical practice questions. References of key documents or websites identified were then further explored for additional information.

Websites of national bodies, societies, research organisations, patient-groups and charities associated with the management of colorectal cancer were also searched with the purpose to identify data sources.

6.3.3 Query of National databases of patient information for data

Online data mining identified a range of potential data sources and the most appropriate data sources were approached for anonymised data. Discussions for potential linkage of multiple specialist databases were also carried out with database specialists in order to establish the feasibility of the data collection for the specific data queries.

6.4 Results

6.4.1 Formulation of data queries (current clinical practice questions)

The following section presents the fifteen guideline topics A-O with a 3-step breakdown of the development process from guideline question to formulated current clinical practice data query for each topic.

Guideline topic A: diagnosing CRC

Guideline question
What is the most effective initial diagnostic intervention(s) for patients with suspected CRC to establish a diagnosis?

↓

Information linking the guideline question to current clinical practice.[142]
The optimum diagnostic strategy for colorectal cancer has not yet been defined. Historically, the interventions used in each centre are guided by local expertise and preference. The aim of the investigation is to achieve adequate examination of the entire colon and rectum. In the past barium enema was the investigation of choice. There is no need for patient sedation and it carries a low incidence of serious complications. The introduction of the highly sensitive fibre-optic endoscopic examinations has seen a huge rise in flexible sigmoidoscopy (endoscopic investigation of the distal 50cm of the large bowel) and colonoscopy (complete endoscopic examination of the rectum and colon). Colonoscopy is considered the gold standard. Some centres may however offer patients a combined investigative pathway of flexible sigmoidoscopy followed by barium enema as an alternative way to image the entire colon. CT colonography (or CT colonoscopy, or virtual colonoscopy) is a recently developed modality that is less invasive and does not require sedation.

↓

Current clinical practice question:
What initial investigation is currently being performed in units across the UK for patients with suspected CRC in order to establish a diagnosis? Are patients first being investigated with: barium enema, flexible sigmoidoscopy, barium enema and flexible sigmoidoscopy, colonoscopy, CT colonography

Guideline topic B: staging CRC

Guideline question:

For patients diagnosed with primary colorectal cancer, what is the most effective technique(s) in order to accurately stage the disease (excluding pathology)?

↓

Information linking the guideline question to current clinical practice. [142]

The optimal modality may vary depending on the clinical situation. Historically, staging of CRC relied on physical examination including digital rectal examination. More recently staging relies mainly on CT, endo-anal ultrasound scan (EUS) and MRI. In addition PET and PETCT have been introduced in the assessment of distant metastases. Availability of MRI, PET, PETCT and EUS in particular may differ from one centre to another and additionally there may be differing levels of local expertise with regard to the interpretation of these images; this may lead to variation in the staging process offered to patients. In addition, modalities may differ in their ability to accurately demonstrate distant metastases, assess early cancers (T1 muscularis propria invasion and wall penetration), define the mesorectal fascia and the circumferential margin (CRM), and to a certain degree predict the suitability for restorative surgery (low anterior resection rather than abdominoperineal resection).

↓

Current clinical practice question(s):

What modality is currently used in the UK to stage colon and rectal cancer?

Do all patients get a CT chest-abdomen-pelvis?

Which group of patients get staged with PET or PETCT?

Which group of patients get staged with MRI?

Which patients with rectal cancer receive EUS as part of their staging?

Guideline topic C: prognostic factors for curative treatment of stage I/polyp cancers

Guideline question:

For patients who have undergone local excision for stage I colorectal cancer, including polyp cancers, with/without neoadjuvant treatment for rectal tumours, which prognostic factors determine the most effective curative treatment?

↓

Information linking the guideline question to current clinical practice.[143]

Patients with Stage I CRC have a five-year survival of >95% following surgery, segmental resection with clear surgical margins (removal of a segment of large bowel including its associated mesentery). Surgery is the curative treatment.

Stage I colorectal cancer may also be identified in endoscopically resected polyps (malignant polyps). Less commonly, it may be found in polypoid lesions resected en-bloc with Endoscopic Submucosal Dissection (ESD) or Transanal Endoscopic Micro Surgery (TEMS).

The UK Bowel Cancer Screening Programme has lead to increased frequency of malignant colonic polyps. Almost all locally removed malignant polyps are Stage I cancers. Endoscopic resection of malignant polyps may be sufficient as the only management but there is a risk of local recurrence or metastatic spread, particularly to local lymph nodes, since the mesentery, which contains the nodes, is not resected. These risks may be reduced by subsequent surgery, but the associated risks such as bleeding, infection or possibly death, and the effects on quality of life need to be balanced against the potential benefits.

↓

Current clinical practice question:

What treatment are patients currently having in the UK for stage I / polyp cancers?

Local treatment (endoscopic or other)?

Surgical segmental resection?

Guideline topic D: self-expanding metal stents (SEMS) for malignant obstruction

Guideline question

For patients with acute large bowel obstruction as a first presentation of CRC

A) Should all patients have a CT scan to confirm diagnosis and stage?

B) What are the indications for stenting as a bridge to elective surgery?

C) What is the optimal timing for stenting to occur?

↓

Information linking the guideline question to current clinical practice.[143]

Up to 30% of CRC cases in the UK present in the emergency setting. Emergency surgery performed for obstructing lesions is associated with a high morbidity and also with peri-operative mortality ranging from 10-20% compared to 5% in the cases of elective cases. In addition, emergency surgery results in a higher rate of stoma formation, high intensive care use and prolonged hospital stay. The introduction of SEMS has provided the opportunity for endoscopic decompression of these patients in an attempt to reduce the risk of emergency surgery. Following decompression it is possible to correct electrolyte imbalance, evaluate the extent of disease, determine the presence of synchronous lesions and evaluate co-morbidities, thus enabling the planning of the most appropriate elective surgery. The placement of a SEMS however is not without risk. It can be associated with colonic perforation, stent migration, malposition or may delay surgery further if the procedure is unsuccessful. The incidence of stent-related complication increases the longer the stent remains in situ. It has been suggested that the success rate for stent insertion is lower for tumours proximal to the sigmoid colon, but with the advent of newer devices able to pass through the endoscopic therapy channel the success of stent placement in the right colon is likely to increase. The potential hazards of SEMS placement must be balanced against the lower surgical mortality in cases of emergency surgery for right-sided colonic obstruction when compared to left-sided lesions.

↓

Current clinical practice question:

Are all patients with a diagnosis of malignant bowel obstruction getting a CT scan?

Who is getting a stent and who is getting emergency surgery?

In how many patients is the bowel currently being defunctioned?

Guideline topic E: pre-operative management for non-metastatic locally advanced colon and rectal tumours

(locally advanced tumours are defined as those tumours that appear unresectable or borderline resectable at presentation)

Guideline question

For patients presenting with a) non metastatic locally advanced colon cancer is pre-operative chemotherapy followed by surgery more effective than immediate surgery and for patients presenting with b) locally advanced rectal cancer is pre-operative radiotherapy, pre-operative chemotherapy or pre-operative chemoradiotherapy more effective than immediate surgery?

↓

Information linking the guideline question to current clinical practice.[143]

Colon cancer occurs at several different sites along the large bowel with variation in the anatomy affected. For most of these sites, the main risk is peritoneal involvement which when it occurs is usually widespread. Any strategy to reduce the risk of recurrence needs to have a systemic approach. However it is not known whether pre-operative chemotherapy is able to reduce the risk of this type of recurrence. Pre-operative chemoradiotherapy is given to patients with locally advanced rectal cancer, with the intention of reducing tumour size to facilitate potentially curative surgery. There is concern that for a small proportion of patients their tumour may progress while on such therapy, thereby losing the window of opportunity for surgical resection. There is also concern that pre-operative chemoradiotherapy is being used for the treatment of very low rectal tumours to facilitate sphincter saving surgery (i.e. a low anterior resection versus an abdominoperineal resection).

↓

Current clinical practice question:

What pre-operative treatment are patients with locally advanced but non-metastatic colon and rectal cancer receiving?

Are patients with colon cancer receiving pre-operative chemotherapy?

Are patients with rectal cancer receiving pre-operative chemoradiotherapy?

Guideline topic F: the most effective sequence of chemotherapy and surgery for the treatment of patients with CRC and synchronous metastatic disease

Guideline question

In patients with CRC presenting with overt synchronous metastatic disease, what is the effectiveness of treating metastatic disease before, after or at the same time as treating the primary tumour?

↓

Information linking the guideline question to current clinical practice.

At presentation approximately 25% of patients with CRC have metastatic disease. In these patients it is thought that the outcome is worse than in those patients that present with metachronous metastatic disease.

The first issue to be addressed is whether the primary tumour is causing obstruction. If this is the case then surgery either to resect or to bypass the tumour should be the primary treatment before considering both chemotherapy and surgery for the metastatic disease. In some cases a stent of the obstructing primary tumour may be possible.

In patients with resectable metastatic disease they should undergo both surgery and systemic treatment with chemotherapy, as their disease is potentially curable. The questions are: chemotherapy prior to or after surgery? Should surgery be staged or combined?

For patients with unresectable disease all treatments are palliative and the main issue is whether leaving the primary tumour in situ is harmful to the patient.

↓

Current clinical practice question

For patients with resectable synchronous metastatic disease are units offering pre- or post-operative chemotherapy or are they offering a combination of both? In these patients is the surgery staged or a combined procedure?

In patients with non-resectable metastatic synchronous disease are units offering adjuvant chemotherapy and are they offering surgery for the primary tumour?

Guideline topic G: neoadjuvant radio- and chemoradiotherapy for rectal cancer

Guideline question:

For patients with operable rectal cancer, what is the effectiveness of short-course pre-operative radiotherapy (SCPRT) and chemoradiotherapy?

↓

Information linking the guideline question to current clinical practice.[143]

SCPRT and chemoradiotherapy are widely used to reduce the risks of local recurrence compared with surgery alone, but there is uncertainty over which schedule to use in which particular clinical setting. SCPRT is a brief (5 days) treatment with high dose per fraction radiotherapy. Short-term side effects are minimal though there is some risk from long-term morbidity. Chemoradiotherapy involves a protracted (minimum of 5 weeks) course of radiotherapy with concomitant chemotherapy. Short-term side effects are more marked and long-term effects can occur. Some cases respond completely.

↓

Current clinical practice questions

Do all patients undergoing an abdominoperineal resection for low rectal cancer receive pre-operative radiotherapy?

For patients allocated to receive pre-operative radiotherapy do they all receive long-course chemoradiotherapy?

If short-course radiotherapy is considered for some or all of these patients under what circumstances is this done?

How is the decision made between short-course radiotherapy versus long-course chemoradiotherapy for patients with rectal cancer requiring an abdominoperineal resection (APR)? Is there a standard policy in the units / networks?

Is a re-staging MRI or other imaging investigation performed after the completion of the pre-operative neoadjuvant treatment?

Is a biopsy of the downstaged tumour site performed to confirm downstaged appearance on imaging?

If the re-staging MRI identifies the original tumour to be completely downstaged to R0 and the biopsy of the site is negative what decision is made then? Do units proceed with performing an APR as originally planned or do they offer regular follow-up with re-imaging and re-biopsy**?**

Guideline topic H: adjuvant chemotherapy for stage II and III rectal cancer.

Guideline question

In patients with clinical or pathological stage II and III rectal cancer what is the effectiveness of adjuvant chemotherapy following surgery?

↓

Information linking the guideline question to current clinical practice.[143]

Colonic and rectal tumours are anatomically in continuity and similar in histopathology. When metastatic, both respond to cytotoxic chemotherapy to a similar level.

Patients are assessed pre-operatively for their risk of recurrence by clinical examination and imaging i.e. clinical staging (cTcNcM). Patients are assessed post-operatively for their risk of recurrence by virtue of the surgical specimen i.e. pathological staging (pTpNpM).

Although it is assumed that the effects of post-operative adjuvant chemotherapy achieved in colon cancer will be the same as in rectal cancer, there has been less direct evidence to support this in patients who have received pre-operative radiotherapy or chemoradiotherapy. Also historical trials in the post-operative setting have used a combination of chemotherapy concurrent with radiotherapy, which has made the assessment of the impact of adjuvant chemotherapy more difficult.

↓

Current clinical practice question:

Are patients with stage II and III colon and rectal cancer being offered adjuvant chemotherapy?

Guideline topic I: adjuvant chemotherapy for high risk stage II colon cancer

Guideline question

For patients with high-risk stage II colon cancer what is the effectiveness of adjuvant chemotherapy after surgery?

↓

Information linking the guideline question to current clinical practice.[143]

The benefit of adjuvant chemotherapy was first demonstrated for patients with stage III disease. In the MOSAIC and NSABP C-07 studies 40% and 29% respectively of patients had stage II disease. The remainder had stage III disease. It is recognised that patients with stage II disease have a better prognosis than those with stage III disease and therefore the benefit of adjuvant chemotherapy is likely to be less. It is known that the prognosis of patients with stage II disease is variable and efforts have been made to identify those at higher risk of relapse.

↓

Current clinical practice questions:

Are patients with high-risk stage II disease being offered adjuvant chemotherapy?

If so how are units defining high risk in patients with stage II disease?

poorly differentiated tumours

extra-mural vascular invasion

T4 tumours (local extension or perforation)

obstructed tumours

small number of lymph nodes harvested (which means the patient has been inadequately staged)

microsatellite instability

mucinous tumours

tumour budding

Guideline topic J: adjuncts for patients with unresectable metastatic disease

Guideline question

What is the most effective additional treatment (adjuncts) to systemic chemotherapy to achieve cure or long-term survival in patients with apparently unresectable metastatic disease?

↓

Information linking the guideline question to current clinical practice.[144]

Where metastatic disease is considered unresectable, systemic combination chemotherapy, with or without biological agents, is the standard of care. Long-term cure is unlikely but median survival can be prolonged to approximately 2 years. Provided a good response is seen in patients with unresectable liver, lung or peritoneal disease following chemotherapy, then local procedures can be attempted to try to prolong the disease-free interval. These local procedures have been most applied to the liver where radiofrequency ablation (RFA) is the most commonly used local treatment, although conclusive data on the benefits have not yet been published. There are even less data on alternative local procedures such as microwave, laser, cryotherapy, radio-embolisation or stereotactic body radiotherapy (SBRT). Some of these local procedures can also be applied to lung metastases, depending on the size and position of individual lesions.

↓

Current clinical practice question:

Are patients with unresectable metastatic disease who have had a positive response to systemic chemotherapy being offered adjunct local treatments to prolong their disease-free interval?

If so which one?

Guideline topic K: imaging of liver metastases

Guideline question

In a patient with colorectal cancer metastasised to the liver which imaging modalities most accurately determine the number and extent of metastases pre-operatively?

↓

Information linking the guideline question to current clinical practice.[144]

In the UK the diagnosis of liver metastases is usually derived from a CT scan performed as part of the original staging or during follow-up after potentially curative surgery for the primary cancer. Currently the two imaging modalities used to assess the presence, number and extent of liver metastases in order to decide whether a lesion is operable are MRI with liver contrast enhancement and PETCT scans. Most patients prepared for surgery will have both of these investigations at some point in their pre-surgical assessment. PETCT is considered by many to be more accurate in detecting liver metastases. However it is an expensive investigation, and not widely available. Patients need to travel long distances to gain access to a PETCT scanner. MRI is much more widely available and has smaller costs.

↓

Current clinical practice question:

What imaging investigation are patients with hepatic metastases offered in their assessment as candidates for curative liver resection?

PETCT

MRI

PETCT and MRI

Guideline topic L: imaging of extra-hepatic metastases

Guideline question

In a patient with colorectal cancer and extra-hepatic metastases (e.g. lung, brain, peritoneum), which imaging modality most accurately determines the extent of metastases?

↓

Information linking the guideline question to current clinical practice.[144]

The common sites of extra-hepatic metastases are distant lymph nodes, peritoneum and lungs. Rare sites of metastases include adrenal glands, central nervous system and bones. Having detected extra-hepatic disease, it is important to determine the extent of disease to offer the appropriate treatment strategy. Information is obtained by means of contrast-enhanced CT scanning of chest, abdomen and pelvis. Further information is also obtained using MRI and PETCT, both for lesion characterisation and also for evaluation of extent and site of extra-hepatic tumour burden. However, little is known as to which is the most useful investigation or the correct sequence of investigations to accurately determine the extent of tumour burden in these patients.

↓

Current clinical practice question:

What imaging investigation are patients with extra-hepatic metastases offered in their assessment as candidates for metastasectomy?

PETCT

MRI

PETCT and MRI

Guideline topic M: chemotherapy regimens for patients with advanced and metastatic colorectal cancer

Guideline question:

What is the effectiveness of oxaliplatin- and irinotecan-based chemotherapy regimens for patients with advanced and metastatic colorectal cancer?

↓

Information linking the guideline question to current clinical practice.[144]

Both oxaliplatin and irinotecan have assumed important roles in the management of colorectal cancer – both in combination with fluoropyrimidines and also, for irinotecan, as a single agent. When combinations of oxaliplatin and a fluoropyrimidine are compared against irinotecan combinations then generally the results are equal, albeit with differing toxicities. Irinotecan appears to have activity both in combination with a fluoropyrimidine and as a single agent. The combination regimens seem to have less toxicity, and appear to demonstrate a trend to better outcomes than when used as a single agent.

Currently, for patients with advanced metastatic disease, both oxaliplatin and irinotecan can be used to extend disease-free and overall survival. There are a number of less frequent circumstances (for example liver-limited metastatic disease), where alternative strategies are used but these are with the intention of long-term disease control, rather than palliation. Defining the optimal strategy for sequencing of these agents remains a difficult management issue.

↓

Current clinical practice question:

What chemotherapy combinations / sequence are patients with metastatic colorectal cancer being offered?

Guideline topic N: follow-up for CRC

Guideline question:

In asymptomatic patients who have undergone treatment with curative intent for CRC, what are the optimal method(s), frequency and duration of follow-up?

↓

Information linking the guideline question to current clinical practice.[145]

Whether systematic follow-up for CRC can alter long term clinical outcome remains controversial. It is also not clear to what extent follow-up can be tailored to the risk of recurrence as defined by pathological stage. A practicing clinician can accumulate a large cohort of follow-up patients and this surveillance can consume significant resources. In addition, what constitutes good clinical practice in terms of follow-up has not been established and there is enormous variation in terms of frequency, duration, clinical setting and interventions employed. Many centres use a policy of CT scanning at variable intervals, with or without serial serum CEA estimation to detect liver and/or lung metastases during the first few years after initial curative resection. Colonoscopy at various time intervals serves the purpose of surveillance for local recurrence of tumour or metachronous tumours. There is also the issue of the effect that follow-up has on quality of life.

↓

Current clinical practice question:

Are all patients with CRC offered follow-up after their treatment?

Is follow-up dependent on pathological stage of disease?

What imaging modality is used for follow-up and how frequently?

What serological tests are offered for follow-up and how frequently?

What endoscopic surveillance is offered and how frequently?

Guideline topic O: information for patients with CRC

Guideline question:

For patients with colorectal cancer, what are the information needs associated with bowel function?

↓

Information linking the guideline question to current clinical practice.[145]

Treatment for colorectal cancer often causes a change in bowel function. This can be distressing for patients and have other adverse effects, including dietary restrictions and changes in body image and sexual function. Patients want to know what to expect after surgery, what is normal and when they should seek further medical advice. Clear and effective communication of information can improve well-being and quality of life. There is paucity of data on this topic from trial data and information that has traditionally been available to patients has been delivered by interested healthcare professionals who have compiled the information based on their own professional experience. However the key question is: ‘what do patients identify as their information needs on the topic of bowel function in relation to CRC?’.

↓

Current clinical practice question:

What information is currently available to patients with a diagnosis of CRC on the topic of bowel function?

Nationally produced leaflet?

Locally produced leaflet?

Access to relevant support websites?

No information

6.4.2 Data sources

With the current clinical practice data queries formulated the online search identified the following national databases as potential sources of answers.

Hospital Episode Statistics (HES)

HES is a records-based system that covers all NHS trusts in England, including acute hospitals, primary care trusts and mental health trusts and contains details of all admissions, outpatient appointments and A & E attendances. This data is collected during a patient's time at hospital and is submitted to allow hospitals to be paid for the care they deliver. [146]

The data collected relate to patients (including age, sex, ethnicity and location of residence), clinical details (including diagnoses, operative procedures, consultant and specialty), administrative details (such as NHS trust, GP, admission and discharge date, source of referral).[146]

Hospital Episode statistics (HES) was originally conceived in 1987 following a report on collection and use of hospital activity information (Korner report). HES aims to collect a detailed record for each hospital ‘episode’ of admitted patient care. [140,141]

For many years clinical practice data for HES was collected from in-patients by the completion of the Korner forms, which were often used as a poor, sometimes illegible, discharge summary for the general practitioner.[141]

Highly skilled clinical coders are required to convert clinical diagnostic terms gleaned from the notes or Korner returns into data entries under the International Classification of Diseases version 10 (ICD-10). Their skills are also required to convert interventions into data entries under the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures version 4.5 (OPCS-4.5) through the Trusts’ Patient Administration Systems to a British Telecom database management system called Secondary User Services (SUS).[141]

Since the introduction of ‘Payment by Results’ (PbR) the terms have also been put through a ‘grouper’ to create the necessary Health Resource Groups for the purposes of reimbursement. Extracts of SUS data are anonymised and cleaned and made available by the NHS Information Centre (NHS-IC) for secondary use as HES data.[141]

As HES data is very complex, any standard report can only show basic information such as number of episodes with a certain primary diagnosis or primary procedure. As might be expected, patients often present with more than one diagnosis or have more than one procedure performed. [146]

The team of professionals at the National Cancer Services Analysis Team (NATCANSAT) use data science to provide analysis based on combinations of procedures and diagnoses as required. Data Science is the application of analytical and computing skills using large quantities of data.[147]

As well as managing large datasets, NATCANSAT also manipulate, link and analyse the data in order to present the information in interesting, innovative and ultimately useful ways. This enables the data to be used in order to answer specific needs or queries.[147]

The Patient Episode Database for Wales (PEDW)

PEDW records all episodes of in-patient and day-case activity in NHS Wales hospitals. This includes planned and emergency admissions, minor and major operations, and hospital stays for giving birth. Hospital activity for Welsh residents treated in hospitals in England is also included.[148]

The data are collected and coded at each hospital. The records are then electronically transferred to the NHS Wales Informatics Service (NWIS) - previously known as Health Solutions Wales (HSW) - where they are validated and merged into the main database.[148]

The National Bowel Cancer Audit Project (NBOCAP)

The overall aim of clinical audits is to improve patient outcomes by improving professional practice and the general quality of the services delivered.[149]

The bowel audit project began in the late 1990's following an approach by the Joint Consultants Committee to several specialist groups including the Association of Coloproctology of Great Britain and Ireland (ACPGBI) regarding a quality control initiative.[150]

The intention was to establish a series of professionally led studies that would define outcomes and benchmarks in specific areas of care. ACPGBI initiated an audit of malignant large bowel obstruction and this paved the way for successive audits of Colorectal Cancer outcomes throughout the UK over the last decade.[150]

In 2000 the ACPGBI published the first edition of the NBOCAP and has since published audit findings annually aiming to improve the quality of care and survival of patients with bowel cancer.[150]

The National Cancer Services Analysis Team (NATCANSAT), which was established in 1996, provided among other services medical informatics services to the NHS, including involvement in cancer clinical audits. The NATCANSAT website initially provided access to the software that enabled the health-care professionals to collect data for the Colorectal Database of the Association of Coloproctology of Great Britain and Ireland.[149]

The national bowel cancer audit (now abbreviated as NBCA) is a collaborative clinical audit run jointly by the NHS Information Centre and the ACPGBI.[151]

The NHS Information Centre manages a number of audit projects in a number of priority areas including cancer. It is the single largest provider of clinical audits. Each audit is backed by appropriate professional bodies, which provide clinical leadership and direction. The aim is to deliver each audit, which includes project management of the complete process, from development of the data requirements, gaining

national authorisations, collection and secure storage of the data, analysis of findings, and continuing review to ensure improvement.[151]

It also provides the infrastructure for the audits, including: providing a secure clinical audit database and technical infrastructure to ensure patients' data is safe and secure; co-ordinating the various approaches to clinical audit across NHS organisations and professional bodies; and ensuring that the clinical data collected is risk adjusted and aligned with national priorities.[151]

For NBOCAP, the NHS Information Centre's National Clinical Audit Support Programme (NCASP) provides the necessary project management and the technical infrastructure described above. It works in collaboration with a range of NHS organisations and professional bodies to provide the infrastructure for the collation, analysis and feedback of local clinical data needed to support effective clinical audits across the NHS.[149] ACPGBI provides clinical leadership and direction for the audit.[150]

The audit includes all NHS Trusts in England and Health Boards in Wales.
All participating trusts submit their data via the Clinical Audit Platform. The Welsh data is submitted directly from the Cancer Network Information System Cymru (CANISC) to the Clinical Audit Platform.[152]

Participation in the audit has always been voluntary but is strongly encouraged by the ACPGBI. Since its inception, there has been constant improvement in the levels of participation from hospital trusts and improvement in case ascertainment and data completeness.[153]

There are 38 essential data items that are collected as part of the audit. These are listed in table 6.1

Table 6.1 The 38 data items of the NBOCAP database.[153]

	NBOCAP Data item	Additional comments
1	Organisation code	
2	NHS number	
3	Local patient identifier	
4	Date of Birth	
5	Height	
6	Weight	
7	Sex	
8	Postcode	
9	Date of diagnosis	
10	Colonoscopy result	Normal-cancer not seen Abnormal-cancer or polyp seen Inadequate-bowel not visualised Not done
11	Colonoscopy incomplete - reason	Obstructing cancer Poor bowel prep Patient intolerance Technical reason other
12	CT result	M0 M1 uncertain
13	MRI result	T stage
14	MRI result	N stage
15	MRI result	Margins threatened – yes / no
16	MDT discussion	Yes / no
17	Seen by specialist nurse	Yes / no
18	Primary cancer site	10 sites appendix to rectum synchronous cancer not recorded anal cancers not recorded
19	Height of tumour above anal verge	

20	Distant metastases	None Present Uncertain
21	Date surgical procedure carried out	
22	Surgical urgency	Elective Scheduled (=expedited) Urgent (in 24h) Emergency (life saving)
23	ASA grade	
24	Surgical access	Open Laparoscopic then open Laparoscopic converted to open Laparoscopic completed
25	Surgical procedure	
26	Radiotherapy treatment (rectal only)	None Neoadjuvant short Neoadjuvant long Adjuvant Definitive with no plan for surgery Palliative
27	Venous invasion	Yes / no
28	Number of lymph nodes examined	
29	Number of lymph nodes found positive	
30	Circumferential margin involved	Yes / no
31	Dukes stage	
32	T stage	
33	N stage	
34	M stage	
35	stoma	Not performed Ileostomy temporary Ileostomy permanent Colostomy temporary Colostomy permanent

36	Date of death	
37	Discharge date	
38	Major surgical complications (The definition of a major complication is one that required re-operation, or, interventional radiology, or ITU/HDU care, or delayed discharge for more than 72 hours.)	None Leak Abscess Bleed Obstruction Stoma – as a second procedure Re-admission within 14 days other

The linked National Cancer Data Repository (NCDR)

The National Cancer Information Network (NCIN) was set up to co-ordinate the collection, analysis and appropriate distribution of patient cancer data by building, maintaining and quality assuring a new national repository of cancer data. [154]

The National cancer data repository (NCDR) contains merged data on all cancer patients in England from the following sources:[83]

8 English Cancer Registry Data (1990-2010)

(This dataset provides details of cancer diagnoses and demographic information about cancer patients.)

ONS Minimum Cancer Dataset (1985-2010)

(Basic information from the Office for National Statistics database of cancer registrations in England, allowing the repository to be reconciled to official national statistics.)

Hospital Episode Statistics (April 1997-March 2010

(Data on inpatient and day case hospital episodes for patients with a diagnosis of cancer in the admitted patient Hospital Episodes Statistics database. This dataset provides clinical information about diagnoses and operations as well as further demographic and administrative details.)

National Clinical Audit Data

(Data from the national bowel, lung, and head & neck cancer clinical audits. These data provide a range of detailed information on diagnosis and treatment of patients for specific cancers).

General Practice Research Database

(An indirect linkage between the NCDR and the General Practice Research Database (GPRD) also exists and applications for linked data may be made to GPRD. GPRD collects data on over 3.6 million active patients (approximately 13 million total) from around 488 primary care practices throughout the UK. This linkage provides vital information on primary care that is not included in the other available datasets.)

The NCDR database is updated yearly. The addition of Celtic Countries NCDR Analysis Dataset is also now included. Each of the English cancer registries holds a copy of the database. Applications for data from the repository may be made in line with NHS data access policies. The details of any request should be discussed with the relevant cancer registry at an early stage.[83]

NCIN seeks to improve collaboration and coordination between UK cancer registries and other organisations involved in collecting and analysing information about cancer patients and to raise national standards of timeliness, comprehensiveness and consistency. NCIN also requires cancer registries to take on national lead roles for specific sites of cancer.[155]

The Northern and Yorkshire Cancer Registry and Information Service (NYCRIS) is playing a significant role within the new National Cancer Intelligence Network (NCIN)

and making use of new sources of information about cancer patients being provided through NCIN e.g. Hospital Episode Statistics (HES).[156]

Within NCIN, NYCRIS has national responsibility for leading on colorectal and haematological cancers. Its purpose is to deliver a comprehensive, electronically acquired and linked, high quality cancer dataset that meets the needs of all its stakeholders. This repository for electronic data will become a holding area for all electronic datasets and will allow assessment of data quality and availability. The linked National Cancer Data Repository aims to provide a more extensive source of patient data through the merging of both clinical practice data through HES and other available patient data through other national databases. NYCRIS is working on linkage and cleaning of HES data.[156]

NYCRIS offers an ad-hoc information request service. This processes an average of 350 information requests each year. NYCRIS also provides information in response to individual requests.[156]

Database of Individual Patient Experience (DIPEX)

The Health Experience Research Group at the University of Oxford with funding from the Department of Health has created a database of personal and patient experiences through in-depth qualitative research into over 50 different illnesses and health conditions including bowel cancer. The results of the research are published on two websites – www.healthtalkonline.org and www.youthhealthtalk.org.

The methodology includes rigorous and systematic research methods to sample, collect and analyse interviews with individuals of all ethnic groups over the age of sixteen. These methods provide a high quality evidence-based approach to patient experience and ensure that a full range of patients' perspectives are analysed in terms of what someone might expect to experience when diagnosed with a particular condition or illness.[157]

Users of the database will find accounts – presented through video, audio and written material – on issues such as reaction to diagnosis, consultation with the

doctor, effect on work, social life and relationships, decisions on treatment options and side-effects of treatments.[157]

NHS patient survey programme

Another possible source of current practice data is national patient surveys particularly for data regarding survivorship, which is so important for patients diagnosed with cancer.

Survivorship focuses on the health and life of a person with cancer post treatment until the end of life. It covers the physical, psychosocial, and economic issues of cancer, beyond the diagnosis and treatment phases. Survivorship includes issues related to the ability to get health care and follow-up treatment, late effects of treatment, second cancers, and quality of life. Family members, friends, and care-givers are also considered part of the survivorship experience.[158]

The National Surveys of NHS Patients programme set up in 1997 comprises a series of surveys designed to contribute to monitoring the performance of the NHS as seen from the patient's perspective. The programme systematically gathers the views of patients about the care they have recently received. The information collected aims to allow systematic comparisons of experiences over time and between different parts of the country.[159]

Under the auspice of the NCIN annual cancer patient experience surveys are conducted collecting a large amount of data on how patients rate their cancer journey.[160]

6.4.3 Data retrieval feasibility assessment

A number of national databases were identified as potential sources of current clinical practice data relating to the colorectal cancer guideline questions.

NATCANSAT for HES data, WCSU for Welsh hospital episode statistics data, and NYCRIS for NCDR access were initially approached for a feasibility assessment.

The clinical practice questions for all topics were presented to the data specialists and their assessment on the feasibility of data retrieval for each question was sought.

NYCRIS as the lead cancer registry for colorectal cancer could provide access to all these databases by performing linkage of databases where appropriate.

The following are the results of the feasibility analysis from NYCRIS.

<u>Topic A – diagnosis of CRC cancer</u>

In order to answer this question patient demographic data needed to be combined with endoscopy and radiology data.

Potential data sources:

The NYCRIS registry database could provide the cancer demographic data. The HES Outpatient episodes database could provide endoscopy dates and radiology dates. The NBOCAP database could also potentially provide information of similar nature.

Initial assessment of feasibility

Project possible due to the NYCRIS/HES/NBOCAP linked database.

Topic B: Staging of CRC cancer.

In order to answer this question patient demographic data needed to be combined with radiology data.

Potential data sources

The NYCRIS registry database could provide the cancer demographic data. The HES Outpatient episodes database could provide radiology dates. The NBOCAP database could also provide information of similar nature.

Initial assessment of feasibility

Project potentially possible due to the NYCRIS/HES/NBOCAP linked database. CT data is known to be of excellent quality. MRI / PET / EUS data has not been assessed in this way and there is uncertainty as to how complete this information is within the databases. The project needs further clarification after a trial data request exercise.

Topic C: prognostic factors for curative treatment of stage I/polyp cancers

In order to answer this question patient demographic data needed to be combined with endoscopy data (if patients had endoscopic removal of their polyps), day case unit data (if patients had local procedures as day cases), and main theatre data (if patients had segmental resections).

Potential data sources

The NYCRIS registry database could provide the cancer demographic data. The HES main database could provide operation dates and the HES – out-patient episodes database could provide endoscopy dates and day case procedure dates. The NBOCAP database could overlap some of this information e.g. the main operation dates.

Initial assessment of feasibility

Project possible due to the NYCRIS/HES/NBOCAP linked database.

Topic D: self-expanding metal stents (SEMS) for malignant obstruction.

To answer this question patient demographic data needs to be combined with radiology data, endoscopy data, and main theatre operative data.

Potential data source

Demographic data on patients with CRC cancer can be provided by the NYCRIS database. CT data with dates can be provided by the HES database. The service of colonic stent insertion in some trusts is provided by radiology departments and in some by endoscopy departments so this data would need to be sought by requesting information both from the HES main and the HES outpatient databases. Operative data can be provided by HES and could also be provided by the NBOCAP database.

Initial assessment of feasibility

Project possible. Good data from the NYCRIS/HES/NBOCAP linked database

Topics E,F,G,H,M Chemotherapy, radiotherapy, histopathology topics

Topic E: Pre-operative management for non-metastatic locally advanced colon and rectal tumours.

Topic F: The most effective sequence of chemotherapy and surgery for the treatment of patients with CRC and synchronous metastatic disease.

Topic G: Neoadjuvant radio- and chemoradiotherapy for rectal cancer.

Topic H: Adjuvant chemotherapy for stage II and III rectal cancer.

Topic I: Adjuvant chemotherapy for high risk stage II colon cancer.

Topic M: Chemotherapy regimens for patients with advanced and metastatic colorectal cancer.

Potential data source

No data immediately available. National databases being formed.

Initial assessment of feasibility. These projects are not currently possible.

Topic J,K,L Hepatic and extrahepatic metastases

Topic J: Adjuncts for patients with unresectable metastatic disease.

Potential data source

No national data available.

Initial assessment of feasibility

Project not possible.

Topic K: Imaging of liver metastases

Topic L: Imaging of extra-hepatic metastases

For the imaging of hepatic and extra hepatic metastases the information request is exclusively radiology data combined with patient demographic data.

Potential data source

NYCRIS/HES/NBOCAP linked database

Initial assessment of feasibility

Project possible.

Topic N: Follow-up for CRC

Potential data source

No data within the national databases. There has been a colorectal cancer audit carried out by Stamatakis et al in 1987 and 1997 in the Wales and Trent regions which showed that huge variation exists in the follow-up protocols followed with no change shown in the decade between the two audits.

Initial assessment of feasibility

Not feasible through national databases.

Topic O: Information for patients with CRC

Potential data source

DIPEX database, National patient survey.

Initial assessment of feasibility

Theoretically possible although it depends whether or not the experiences documented in these databases cover the topic of information required by patients.

6.4.4 Data actually recovered for each formulated clinical practice question

No relevant data was recovered for any of the formulated clinical practice questions.

6.5 Discussion

One of the findings of this study was that of subtle mismatch between the clinical question and the clinical practice question. A good example is topic C where the guideline question focuses on prognostic factors and the clinical practice question focuses on types of treatment for stage I / polyp cancers.

This is not a failing of understanding or an accidental deviation from the question the guideline aims to address by the author.

The clinical practice question was developed by the author gradually after research, and discussion with the GDG members on a particular topic. It aims to highlight what is currently a genuine practice question.

The fact that it deviates in some topics from the actual clinical question the guideline is seeking to answer raises the concern that perhaps in some instances the questions set early on in the guideline development process are not the most clinically relevant.

On a number of occasions after the questions of the guideline had been set some GDG members felt that the questions needed revision. However the NICE methodology on question setting was inflexible in what could be amended after a certain time period had elapsed.

The guideline development process is long and timeframes are tight for the development of each topic. Therefore changes to the original questions are not permitted after the first few stages of the development process.

Though this is understandable it raises the concern that in some cases resources and time are devoted to answering the wrong question.

A more flexible methodology with regard to the formulation of the guideline questions might be a future consideration by NICE.

Alternatively, a lengthier time period devoted to scoping the guideline questions even before the guideline evidence searching begins might avoid such mismatch in the future.

In addition, the results show that for the purpose of a needs assessment report supporting a national clinical guideline the use of current practice data collected in national databases in the UK was not straightforward. This was despite initial indication through the online data mining search of a number of data sources showing promising potential for providing the answers to the formulated queries.

National databases have the potential to be a major source of current practice data, particularly if the information from multiple databases can be linked in order to provide concise answers to particular questions. The HES and PEDW national databases showed initial potential in providing answers to some of the data queries.

Successive audit reports have indicated considerable problems with HES data completeness, accuracy of clinical coding and engagement of clinicians [161,162] even following the introduction of PbR.[163] Generally speaking, clinical coders are well trained and very accurate in converting clinical terms into codes.[164] The problem is that it is difficult for them to extract the correct information from unstructured clinical notes.[141] If HES data were to be used then the issue of the data quality would have to be addressed.

The National Bowel Cancer Audit was another potential source of current practice data relevant for the needs assessment. Though voluntary in its participation the audit has high participation rate from all the hospitals and year on year the data has had steady improvement in its completeness.

However, there are challenges that remain with the data quality and one such issue is that of handling of missing data. The audit data set does not allow the distinction between patients who have not undergone a surgical procedure and those for whom the data item is missing. Siimilar issues arise for diagnostic and staging procedures.

Multiple imputation is used to fill in any missing information but this remains a significant challenge for the national bowel cancer audit and other similar databases.

National audits are designed to measure the quality of patient care and improvements over time. Healthcare professionals and organisations can use the

information contained in the audits to look at the national picture and identify areas for improvement as well as to learn from best practice.[165]

A number of the data items of the national bowel cancer audit could potentially provide answers to some of the formulated data queries but would need to be combined with data from HES and PEDW in order to give more complete answers.

It was clear that the variety of the data from the different databases and the issues of data quality made the project of answering the formulated data queries that much more complex. The data would require careful analysis to ensure quality standards were met.

For these reasons the most exciting and promising finding was that NYCRIS, the lead registry for colorectal cancer was not only an official access point for all these databases but was additionally performing groundbreaking work in the improvement of the quality of the data as well as attempting linkage analyses of data from different databases in order to provide answers that were important clinically.

The NCDR which links all these and other databases together provides exciting opportunities for novel hypothesis-led studies, as well as the possibility of enhancing other datasets. NYCRIS has a leading role in the development of the NCDR.

The unfortunate and rather disappointing result that NYCRIS was not able to deliver any of the planned analyses was due to the timing of the specific data request which came at a time of massive change for NYCRIS and all the cancer registries. Unfortunately there was a lack of resources and the initial enthusiasm for addressing the needs of the project did not lead to results as expected.

Over recent years, cancer registries have been subject to a number of outside influences that have caused them to adapt the way they are organised and to revise their working practices. The roles of registries have developed from traditional registration into cancer information and intelligence provision, with registries having an increasingly important role to play in taking forward the national cancer agenda.

NYCRIS, in recognising the task ahead set out a long-term development plan over three years 2008-2011. This plan encompassed all developments across the organisation including the upgrade of electronic data processing, changes in the registration process, an increase in the information and research outputs, and an increase in staff resources.[155]

The time period of this project overlapped with this period of intense change at NYCRIS. In addition linkage analyses of national databases as a research tool are in their infancy and require time, staff and funding.

Therefore for topics that could have potentially been answered with data from the HES, PEDW and NBOCAP linked databases the conclusion from these results is that the data is collected at a national level, attempts are currently being made to address and to improve the quality of the data but access to the data via NYCRIS at present also needs to be improved.

For radiotherapy and chemotherapy related queries the results show that data collection was even less advanced.

An important finding is the inability to access radiotherapy data at a national level at all. Radiotherapy is a major modality in the treatment of cancer and also represents a significant sector within the NHS, in terms of both workforce and capital investment. Information on radiotherapy activity is recorded in various ways by different radiotherapy departments, with no nationally agreed dataset or data return.[154]

This has since been addressed through an initiative of the NCIN in partnership with the NHS information centre. The development of the radiotherapy dataset (RTDS) allows for the routine collection of clinically and managerially relevant activity data from radiotherapy facilities with good quality reporting, in order to commission or monitor radiotherapy services in an evidence-based manner. The data is collected by the staff at NATCANSAT on behalf of the NCIN.[166,167]

Prior to the inception of the RTDS, very limited radiotherapy data were collected, and

there were a wide variety of definitions of each of the currencies in use. The RTDS seeks to standardise these currencies, and to introduce new currencies which are aligned with other activities in the NHS. [166]

Despite the recent lack of central information on radiotherapy, each patient's data is stored on a database linked to the radiotherapy treatment machine (these systems are referred to generically as Oncology Management Systems), which generate an essential clinical record of the radiation delivery to each patient. These systems have been in use for several years. [166]

A decision was taken early in the RTDS development to use these systems as the main source for the dataset, in order to avoid duplication of effort in entering the radiotherapy treatment details onto hospital patient administration systems (PAS), and to benefit from the excellent data quality in the technical radiotherapy data resulting from the use of the system which actually controls patient treatment.[166]

Another important result finding is the absence of a national chemotherapy database. Information on chemotherapy delivery is rudimentary, largely because some providers of chemotherapy services are still using paper-based systems to prescribe and record activity.[154] This has since been addressed with the development of the Systemic Anti-Cancer Therapy (SACT) Data Standard and dataset as an initiative by NCIN and the NHS Information Centre (NHS IC).[168]

Data on the treatment of patients with chemotherapy will be managed by the Systemic Anti-Cancer Therapy (SACT) Data Standard. With the advent of electronic prescribing systems it is possible to record this complex information on patient management in a standardised way. The SACT data standard defines terms that are used in chemotherapy prescribing for individual patients. It also specifies a reporting dataset.[168]

The standard covers all patients receiving cancer chemotherapy in or funded by the NHS in England. The data standard relates to all cancer patients, both adult and paediatric, in acute in-patient, day-case, out-patient settings and delivery in the

community. It covers chemotherapy treatment for all solid tumour and haematological malignancies, including those in clinical trials.[168]

In the clinical setting its primary use will be in prescribing and administering chemotherapy. The reporting dataset will be used at both local and national levels to generate ‘secondary uses’ information to provide quality data to support service development and commissioning.[168]

In addition the results show that the need for survivorship data was not met by the databases and survey data currently available. Though there has been steady expansion and improvement of the type of data collected and the collection process itself through initiatives of the NCIN this field of clinical data has been chronically underdeveloped and therefore requires more time for data to be collected.

Since the inception of HES the NHS has changed considerably, requirements for data have changed, monitoring of service and outcomes has become a high priority.

The Cancer reform strategy established the National Cancer Intelligence Network (NCIN), under the umbrella of the National Cancer Research Institute (NCRI), with objectives, inter alia, of nationally promoting efficient and effective data collection throughout the cancer patient journey in order to secure improvements in standards of cancer care and clinical outcomes. This also provided the opportunity for the development of appropriate regulations for data access and data processing.[154]

The NCDR, the radiotherapy dataset, the chemotherapy dataset are examples of some of the work already carried out. These are new developments and will take time to assess in terms of data quality but they are important steps in the improvement of national cancer data collection.

Further work is ongoing. The NCIN is working together with the NHS information centre (NHS-IC) to develop a new National Cancer Dataset. This will replace the current National Cancer Dataset and will include both the Cancer Registration dataset and additional site specific data items relevant to the different tumour types.

It will be aligned with the other mandated national cancer datasets (Cancer Waits and Radiotherapy) and with the Systemic Anti Cancer Therapy dataset (SACT).[169]

The new dataset, which is called the Cancer Outcomes and Services Dataset (COSD), will support the current needs of the NHS to provide information on incidence, mortality and survival and also service and outcomes. The intention is to collect data already used for patient management and clinical care and which where possible should mostly be available from existing NHS electronic systems such as PAS, pathology and MDT systems. These data will then be sent to the regional cancer registries who will link these and other multiple data sources at patient level using the NHS number to complete the full dataset.[169]

The 12 NCIN Site-Specific Clinical Reference Groups (SSCRGs) have been engaged to ensure that good quality cancer data is available to inform and to promote improvements in standards of cancer care and clinical outcomes, as well as enabling the use of cancer information to support audit and research programmes.[169]

The English regional cancer registries are involved with the work of the SSCRGs by taking up cancer registration lead roles. Initial work undertaken by each lead registry provided a site-specific "baseline assessment". Registries are now developing their roles as cancer registration site-specific leads through dedicated programmes of work.[170]

At present there is a huge expansion in national database systems. Funding, human resources and technical support are being diverted to this chronically neglected field of health management and this is a welcome improvement.

Although there has been significant multi-directional progress in the national collection mechanism for this type of data in recent years what these results show is that this progress was insufficient and in its infancy with regard to radiotherapy, chemotherapy and survivorship when the present research was taking place and therefore the required data collection for the guideline needs assessment report was not feasible.

However the formulated queries are important as they are attempting to capture the current picture in the delivery of colorectal cancer care and to support a national guideline in its attempt to produce recommendations which can actually be implemented.

This will be considerably obstructed if that current picture cannot be presented. It is absolutely necessary to the implementation of the guideline to know what baseline practice there is. It is impossible for any cancer centre, unit or network to know how well it is doing on delivering care to patients without such key information. Attempting to collect a small sample of this information by means of questionnaire studies has a low return and small local audits and studies are underpowered and unable to provide strong evidence. This information needs to be collected at national level.

Though the absence of a result is disappointing, it is an important finding in itself. Though requests for linkage analyses may come from a variety of authorised sources for the development of policy or for the purposes of commissioning, a request from a guideline development group is an additional viewpoint regarding the needs for health informatics in the NHS. It will hopefully strengthen the plea for faster improvements to be made in this area.

The systematic process used in health needs assessment reports provides the ideal opportunity for communication across professional disciplines and patient groups, gathering evidence about the target patient population, and making use of an evidence-based approach to effect health management changes and improvements.

The value of a health needs assessment lies in the contribution it can make to improving data quality. This is important in meeting quality indicators, in developing disease registers, and in providing information for an evidence-base of need which can support enhanced service provision by creating the basis from which resource and service provision assessments can begin and from which the guideline programme can be developed and implemented.

In addition, the analysis of data which reflects current practice in a variety of ways is in itself a key capability for motivating improvements in quality. Only when the databases are challenged to provide a variety of answers will it become clear whether they can demonstrate appropriate standards of data quality and are adequate for the intended purposes.

Patient data will only be effective if it is collected, analysed and presented in ways that are useful to patients, commissioners, service providers and other interested parties.

There is a responsibility to patients to ensure that their data, which is routinely collected and stored, is appropriately accessible for research and aids the provision of answers to important clinical questions that could help improve their overall management.

Perhaps the unanswered data request of the colorectal guideline needs assessment report can serve as an example of one of the ways in which current-practice data that is collected in the UK might be used with advantage in the future.

7.0 The availability and quality of diagnostic data that make up the evidence to support the most effective method for diagnosing liver metastases from colorectal cancer to assess resectability

7.1 Introduction

The field of diagnostic tests is highly dynamic. New tests are developed at a fast rate and the technology of existing tests is continuously being improved.[171-174]

Exaggerated and biased results from poorly designed and reported diagnostic studies can trigger their premature dissemination and lead physicians into making incorrect treatment decisions.[171-174]

A rigorous evaluation process could not only reduce the number of unwanted clinical consequences relating to misleading estimates of test accuracy, but also limit health care costs by preventing unecessary testing.[171-174]

In studies of diagnostic accuracy, the outcomes from one or more index tests (IT) under evaluation are compared with outcomes from the reference standard (RS), both measured in subjects who are suspected of having the condition of interest.[171,175-177]

The term ‘index test’ (IT) refers to any method for obtaining additional information on the health status of a patient. It includes information from history and clinical examination, laboratory tests, imaging tests, function tests, histopathology.[171,175-177]

The condition of interest or target condition can refer to a particular disease or any other identifiable condition that may prompt clinical actions, such as

further diagnostic testing or the initiation, modification, or termination of treatment. [171,175-177]

In this framework, the reference standard (RS) is considered to be the best available method for establishing the presence or absence of the condition of interest.[171,175-177]

The reference standard can be a single method, or a combination of methods, to establish the presence of the target condition. It can include laboratory tests, imaging tests, and pathology, but also the dedicated clinical follow-up of patients.[171,175-177]

The term accuracy refers to the amount of agreement between the information from the test under evaluation, referred to as the ‘index test’, and the reference standard. [171,175-177]

Diagnostic accuracy can be expressed in many ways including sensitivity and specificity, likelihood ratios, diagnostic odds ratio, and the area under a receiver-operator (ROC) characteristic curve.[171,175-177]

There are several potential threats to the internal and external validity of a study of diagnostic accuracy. The QUality Assessment of Diagnostic Accuracy Studies tool (QUADAS) has been systematically developed to try to guide the assessment of diagnostic studies.[178]

The original QUADAS tool was developed through a collaborative project between the Centre for Reviews and Dissemination of the University of York, and the Academic Medical Centre at the University of Amsterdam. It was funded through the Health Technology Assessment (HTA) programme and was published in 2003.[178]

Since its development QUADAS has been used in a large number of systematic reviews. A modified version of QUADAS, with items related to the quality of reporting removed, has been adopted for use by the Cochrane

Collaboration and is recommended for use in all Cochrane reviews. QUADAS has also been recommended for use by NICE in the assessment of diagnostic data.[178]

In current UK oncological practice the diagnosis of liver metastases is determined from a CT scan performed as part of the original staging or during follow-up after potentially curative surgery for the primary cancer.

Currently the two imaging modalities used to assess the presence, number and extent of liver metastases in order to decide whether a lesion is operable are MRI with liver contrast enhancement and PETCT scans.

Most patients assessed for surgery will have both of these tests at some point in their pre-surgical preparation. PETCT is considered by many to be more accurate in detecting liver metastases. However it is an expensive investigation and not widely available. Patients need to travel long distances to get to a PETCT scanner. MRI is much more widely available and has smaller costs.

7.2 Aim

To research the availability and quality of the diagnostic data that make up the evidence to support the most effective method for diagnosing liver metastases from colorectal cancer to assess resectability.

7.3 Methods

A systematic review and meta-analysis was performed by one reviewer (the author). Training in systematic review methodology was provided by the department of evidence based healthcare at Oxford University and the review was performed according to NICE methodology. The author selected the relevant titles and abstracts from the results of the database search, decided which studies met the inclusion criteria,graded their methodological quality,

extracted the relevant data and entered it into the REVMAN 5 software package for analysis.

7.3.1 The PICO

The traditional PICO framework (patient, intervention, comparison and outcome) used for researching review questions is slightly different for diagnostic topics compared with the more traditional interventional / therapeutic topics.

The PICO for a diagnostic topic maintains the population (P) and outcome (O) items but includes the index test(s) (I) whose diagnostic accuracy is being investigated instead of the intervention more traditionally included in a therapeutic PICO. It also includes the reference standard used to prove the accuracy of the index test instead of the traditional comparator (C) item.

Table 7.1: The PICO for the systematic review on CRC liver metastases

Population	Index Tests (instead of intervention)	Reference standard (instead of comparator)	Outcome
Patients with colorectal cancer	MRI (with liver contrast agent)	Histology of the resected specimen	Sensitivity
AND	PETCT	Follow-up with imaging of those lesions that were not operated	Specificity
Potentially operable liver metastases	CT(with contrast)	Manual palpation of the liver	Survival
	Laparoscopic ultrasound scan	Intra-operative ultrasound scan of the liver	Change in management

<u>The Population</u>

The patients in the studies should be adults (>18 years) of both sexes with known metastatic colorectal neoplasm who are candidates for liver resection. In practice this most commonly means that patients have had a diagnosis of colorectal cancer made on histological biopsy from colonoscopy and they are

potential candidates for liver resection based on the findings of a CT chest/abdomen/pelvis.

Ideally this should be mentioned specifically in the methods of the study. However if studies do not mention specifically how the diagnosis of the colorectal cancer and the liver metastases is made they would not be excluded from the review.

Synchronous and metachronous metastases will be included as well as patients who have had neoadjuvant chemotherapy.

The studies should exclude patients who have contraindication to the contrast media used or to the imaging modalities e.g. pacemakers for MRI or those with advanced liver cirrhosis, or any contraindication to a surgical resection.

The Index test

Studies using all types of liver contrast-enhanced MRI were included. Only studies of fusion PET-CT technology were included.

Studies comparing PET or PET followed by a CT to CT or to MRI were included but only the CT and/or MRI results were entered into the meta-analysis.

No exclusion of any types of CT scanners or laparoscopic ultrasound devices were made on technological grounds.

Reference Standard

There are multiple reference standards, as not all patients will proceed to surgery after their index test is performed.

For the patients who have no resection the reference standard is follow-up imaging.

What defines malignancy or benign pathology on the follow-up scan needs to be stated clearly in the studies.

Usually for malignancy this is defined as any sign of progression in the size of a lesion or regression in size as response to treatment. For benign pathology the definition is no change in the size of a lesion on the scan.

It is important that the period of follow-up is also defined and clearly stated in the studies. 3-6 months is acceptable clinically.

A period any shorter than this might allow insufficient time for progression of a lesion which as a result might be undetectable at follow-up.

A period any longer might result in the follow-up scan identifying a new metastasis rather than the one under surveillance.

For those patients who proceed to liver resection the gold standard reference test is histological analysis of the resected specimen.

In some cases patients will have a combination of manual palpation of the liver, intra-operative ultrasound scan and the histology of the resected specimen.

Some patients who proceed to surgery may prove to have in-operable disease at laparotomy such that no resection is undertaken.

For these patients the reference standard will be manual palpation with or without intra-operative ultrasound scan followed by scanning in the post-operative period if appropriate.

Outcome

The primary outcome was the sensitivity together with the positive predictive value that can be derived from it. The sensitivity is of primary importance as this determines whether metastatic malignant disease is detected or missed (true positives / true positives + false positives).

This was supplemented by specificity reporting to see how balanced the index test is in correctly identifying the patients with negative results and not committing patients to laparotomy unnecessarily (true negatives / true negatives + false positives). Positive predictive value and accuracy values were also calculated.

Secondary outcomes were overall survival and change in management. Although not traditionally associated with diagnostic accuracy studies they are clearly important in clinical decision-making and are necessary in performing cost analyses and economic modelling.

7.3.2 Search

Table 7.2 lists the databases that were searched for this PICO and Table 7.3 lists the truncation symbols that were incorporated into the search. The author closely observed the NCC-C information specialist create and perform the detailed search having received training in the creation of search strategies through the department of evidence based healthcare at oxford university.

In addition hand searching of references in the selected study papers was performed by the author.

The start date for the search was 1995 since advice from experts in radiology indicated that imaging prior to this date produced results, which would be inconsistent with the quality of modern scanning technology.

There was no language filter applied; however translation of foreign language papers was only requested if the abstract was deemed of critical importance to the review. No translations were required for this review.

Table 7.2: List of databases searched for the systematic review on colorectal liver metastases with search dates

Database name	Dates Covered	References retrieved	Finish date of search
Medline	1995-2009	108	01/10/09
Premedline	1995 -2009	21	01/10/09
Embase	1995-2009	95	07/10/09
Cochrane Library	1995-2009	23	21/09/09
Cinahl	1995-2009	1	13/10/09
BNI	1995-2009	0	01/10/09
Psychinfo	1995-2009	0	01/10/09
Web of Science (SCI & SSCI) and ISI Proceedings	1995-2009	128	13/10/090
Biomed Central	1995-2009	1	07/10/09
Total References retrieved (after de-duplication)		287	

The search strategy was as follows:

Colorectal cancer AND Liver mets AND Imaging

1. exp colorectal neoplasms/
2. ((colorect$ or colo rect$) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$)).tw.
3. ((colon or colonic) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$)).tw.
4. ((rectal$ or rectum$) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$)).tw.
5. 1 or 2 or 3 or 4
6. exp Liver Neoplasms/

7. (liver adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$)).tw.
8. exp Neoplasm Metastasis/
9. liver metastas*.tw.
10. hepatic metastas*.tw.
11. hepatic lesion*.tw.
12. exp Neoplasm Recurrence, Local/
13. 8 or 6 or 11 or 7 or 10 or 9 or 12
14. 13 and 5
15. exp Diagnostic Imaging/
16. exp Tomography, X-Ray Computed/
17. exp Magnetic Resonance Imaging/
18. exp Ultrasonography/
19. imaging modalit*.tw.
20. (contrast enhanced CT* or CT*).tw.
21. exp Positron-Emission Tomography/
22. PET*.tw.
23. contrast enhanced MR*.tw.
24. (CT adj (helic* or spiral*)).tw.
25. PET-CT*.tw.
26. exp Neoplasm Staging/
27. exp "Sensitivity and Specificity"/
28. 27 or 25 or 21 or 26 or 17 or 20 or 15 or 22 or 18 or 24 or 16 or 19 or 23
29. 28 and 14

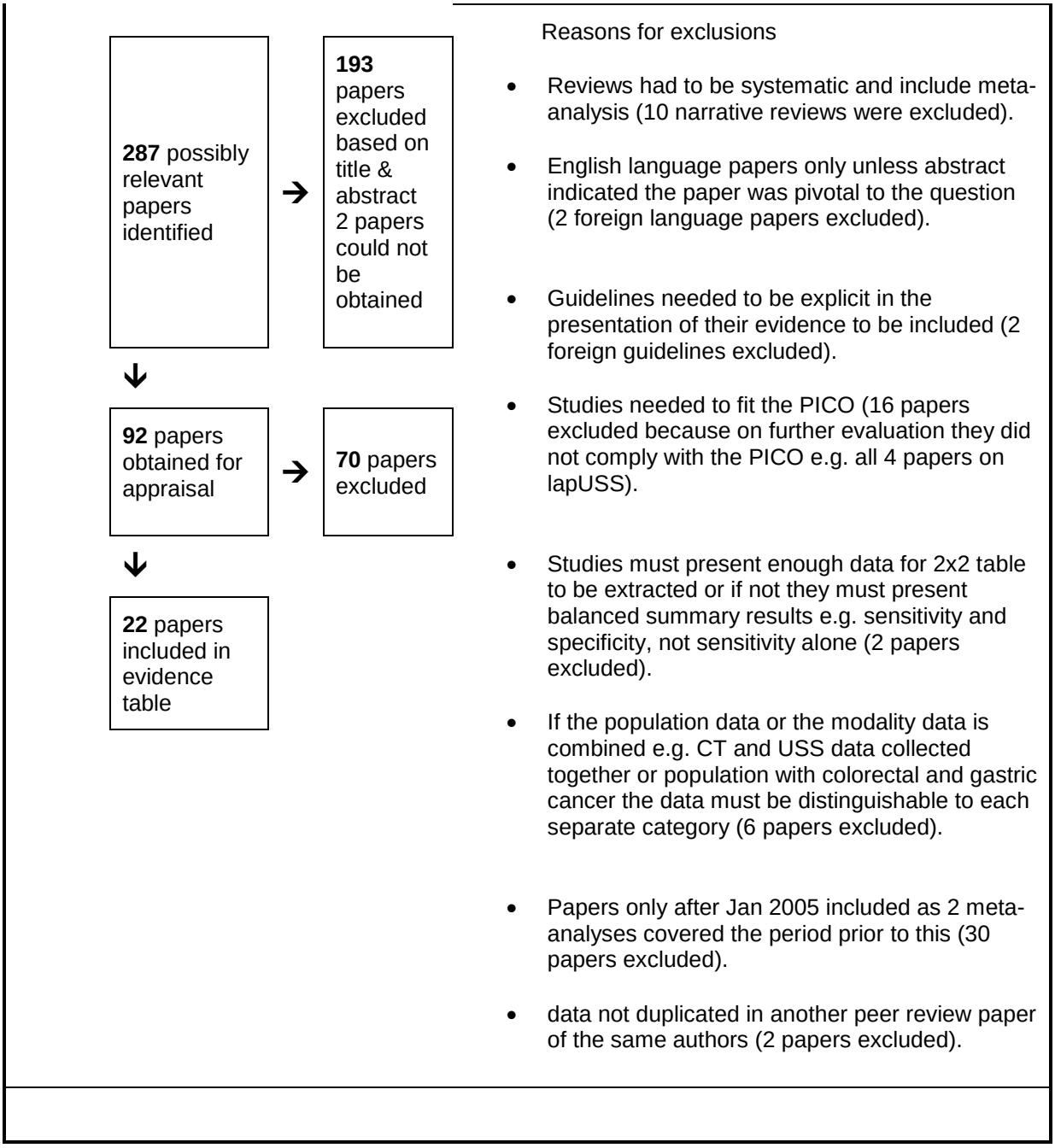
Table 7.3: The truncation symbols incorporated into the search for colorectal liver metastases.[179]

Symbol	Definition
*	All words beginning with a particular stem
$	For all words beginning with a particular stem
?	Overcoming spelling differences and searching for singular and plurals
AND	For all the words
OR	Searches for at least one of the words in the search string
ADJ	For words next to each other in the order specified
()	For words in brackets first / For each word and mapping to MeSH
“	For words next to each other in the order specified

7.3.3 Study selection

Study selection followed the steps that are displayed in Figure 7.1.

Figure 7.1: Flow chart showing the selection criteria for included evidence for the systematic review on colorectal liver metastases



7.3.4 Critical appraisal of included studies

The QUADAS quality assessment tool was used to assess relevant study design characteristics of each study. This was in the form of a NICE

methodology checklist [180] adapted from the original paper [181] and is presented in Appendix 2.

The questions in the checklist are aimed at establishing the validity of the study under review. That is, making sure that it has been carried out carefully, and that the conclusions represent an unbiased assessment of the accuracy and reliability of the test being evaluated. Each question covers an aspect of methodology that is thought to make a difference to the reliability of a study.[180]

Checklist items are worded so that a ‘yes’ response always indicates that the study has been designed and conducted in such a way as to minimise the risk of bias for that item. An ‘unclear’ response to a question may arise when the answer to an item is not reported, or not reported clearly. ‘N/A’ is used when a study of diagnostic test accuracy cannot give an answer of ‘yes’ no matter how well it has been done.[180]

The checklist consists of 14 items and these are listed below [180].

1. Was the spectrum of participants representative of the patients that will receive the test in practice? If it is not there is a potential for spectrum bias.

2. Are selection criteria clearly described? If they are not then there is a potential for selection bias.

3. Does the reference standard classify the target condition correctly? The use of an inappropriate reference standard can bias estimation of the diagnostic accuracy of the index test.

4. Is the time interval between the reference standard and the index test short enough so that there is no change in the condition? If not then there is the potential for disease progression bias of the accuracy results.

5. Was verification with the reference standard given to the whole sample or a random selection? If not then there is the potential for partial verification bias (also known as work-up bias, [primary] selection bias or sequential ordering bias).

6. Did participants receive the same reference standard regardless of the index test result? If not then there is the potential for differential verification bias.

7. Was the reference standard independent of the index test (was not part of it)? When the result of the index test is used in establishing the final diagnosis, incorporation bias may occur. This item will only apply when a composite reference standard is used to verify disease status.

8. Was the index test described clearly? Variation in measures of diagnostic accuracy can sometimes be traced back to differences in the execution of index tests.

9. Was the reference standard described clearly? Variation in measures of diagnostic accuracy can sometimes be traced back to differences in the execution of the reference standards.

10. Was the result of the index test interpreted without knowing the result of the reference standard? This issue is similar to the ‘blinding’ of the people who assess outcomes in intervention studies.

11. Was the result of the reference standard interpreted without knowing the result of the index test? This issue is also similar to the ‘blinding’ of the people who assess outcomes in intervention studies.

12. Was the same data available as in practice? The availability of information on clinical data during the interpretation of test results may affect estimates of test performance. If clinical data will be available when the test is interpreted in practice, then these should also be

available when the test is evaluated. However, if the index test is intended to replace other clinical tests, then clinical data should not be available.

13. Were indeterminate results reported? A diagnostic test can produce these with varying frequency. If these are not reported this may lead to the biased assessment of the test characteristics. Whether bias will arise depends on the possible correlation between indeterminate test results and the true disease status. If indeterminate results occur randomly and are not related to the true disease status of the individual then, in theory, these should not have any effect on test performance.

14. Were withdrawals from the study explained? If participants lost to follow-up differ systematically from those who remain, for whatever reason, then estimates of test performance may be biased.

The above items of the QUADAS checklist can be grouped into four domains. These are patient selection, index test, reference standard, flow (of patients through the study) and timing (of the index test and reference standard). [182]

QUADAS should not be used to generate a summary “quality score”. If a study is judged as “low” on all domains relating to bias then it is appropriate to have an overall judgment of “low risk of bias” for that study. If a study is judged "high" or "unclear" on one or more domains then it may be judged “at risk of bias” or as having “concerns regarding applicability”. [182]

A summary of the results of the QUADAS assessment is advised for all included studies. This could include summarising the number of studies that found low, high or unclear risk of bias/concerns regarding applicability for each domain. If studies are found consistently to rate well or poorly on particular items then reviewers may choose to highlight these.[182]

7.4 Results

The data was extracted individually from all 22 studies included in the review and are presented in detail in the evidence tables of Appendix 3.

7.4.1 Quality Assessment

Design of the included studies

From the 22 studies included in the review [183-204] none were systematic reviews of randomised controlled studies or other study designs.

There were two randomised controlled trials [191,199], and twenty case series.

There were 15 prospective (P)(184-7,190,192,194,196-8,200-4), and 5 retrospective (R)(183,188-9,193,195) case series.

QUADAS quality assessment applied individually to the 22 studies

Table 7.4 lists all 22 studies and how they were assessed for each of the QUADAS items.

In all the studies (100%) the patients were appropriately selected and the selection criteria were clearly described.

All studies (100%) described clearly the index test that was used.

All studies (100%) had multiple reference tests.

These were appropriate for the classification of the target condition, they were applied to the whole study group rather than a random selection, they were not part of the index test, a different reference test was used depending on

the result of the index test, but the same reference test was used for each group of index test results.

However, 45.5% of studies did not report how the reference test was performed.

55% of studies did not report the time between index and reference test.

A large proportion of studies did not report on ‘blinding’.

82% did not report whether the reference test results were interpreted without knowledge of the index test results and 41% did not report on whether the index test results were interpreted without knowing the results of the reference test.

82% of studies did not report their indeterminate results.

73% of studies did not report their withdrawals from the study.

Error!

Study	Were patients representative	Selection criteria clearly described	RS classifies target condition correctly	Time between RS and IT short enough so no change	Was verification with the RS given to whole sample or randon selection	Same RS independent of the IT result	Was the RS independent of the IT - was not part of it	IT described clearly	RS described clearly	IT results interpreted without knowing the result of the RS	RS results interpreted without knowing the reuslts of the IT	Same data available as in practice	Indeterminate results reported	Withdrawals explained	design
Akiyoshi[183]	y	y	y	y	y	n	y	y	n	u	u	y	n	n	R
Arulampalam[184]	y	y	y	u	y	n	y	y	y	u	u	y	n	y	P
Ashraf [185]	y	y	y	u	y	n	y	y	n	u	u	y	n	y	P
Bartolozzi[186]	y	y	y	u	y	n	y	y	y	y	y	y	y	y	P
Bhattacharjya[187]	y	y	y	y	y	n	y	y	y	y	y	y	y	n	P
Cantwell[188]	y	y	y	y	y	n	y	y	n	u	u	y	n	n	R
Chua[189]	y	y	y	u	y	n	y	y	n	n	n	y	n	n	R
Coenegrachts[190]	y	y	y	y	y	n	y	y	y	y	u	y	n	n	P
Kim[191]	y	y	y	y	y	n	y	y	y	y	y	y	y	y	RCT
Koh[192]	y	y	y	y	y	n	y	y	n	y	u	y	n	n	P
Kong[193]	y	y	y	u	y	n	y	y	n	u	u	y	n	n	R
Liu[194]	y	y	y	u	y	n	y	y	y	u	u	y	n	n	P
Nanashima[195]	y	y	y	u	y	n	y	y	n	u	u	y	n	n	R
Orlacchio[196]	y	y	y	u	y	n	y	y	n	y	y	y	n	n	P
Rappeport[197]	y	y	y	y	y	n	y	y	y	y	u	y	n	n	P
Regge[198]	y	y	y	u	y	n	y	y	y	y	u	y	n	n	P
Ruers[199]	y	y	y	y	y	n	y	y	y	y	u	y	n	n	RCT
Schwartz[200]	y	y	y	y	y	n	y	y	y	y	u	y	n	y	P
Selzner[201]	y	y	y	u	y	n	y	y	n	y	u	y	n	n	P
Truant[202]	y	y	y	y	y	n	y	y	y	y	u	y	y	n	P
Vidiri[203]	y	y	y	u	y	n	y	y	n	u	u	y	n	y	P
Wiering[204]	y	y	y	u	y	n	y	y	y	y	n	y	n	n	P

Table 7.4: Summary QUADAS quality analysis of the 22 studies in the systematic review for colorectal liver metastases
RS: reference standard, IT: index test

7.4.2 Meta-analysis

Per patient analysis

12 studies reported CT data per patient, 9 studies reported MRI data per patient, 7 studies reported PETCT data per patient.

CT data

The sensitivity of CT ranged from 47% to 100%. The positive predictive value (PPV) for CT ranged from 86%-100%. Specificity for CT ranged from 0 to 100%. The accuracy for CT ranged from 50% to 98%.

Though there has been no weighting to the following summary values the overall sensitivity and PPV for CT from the 12 studies as calculated from a summary 2x2 table is presented in Table 7.5.

Table 7.5: CT per patient summary values and 2x2 table values

SENSITIVITY		87%		770 / 882
PPV		95%		770 / 770+41
ACCURACY		87%		770 + 266 / 1189
Total TP=770	Total FP=41	Total FN=112	Total TN=266	Total =1189

TP: total positive results; FP: false positive results; FN: false negative results; TN: total negative results; Total: total number of patients investigated by CT

MRI data

The sensitivity of MRI ranged from 50% to 100%. Specificity ranged from 0% to 100%. In a number of studies specificity estimates are not possible as there were no benign lesions identified at all in the population. PPV ranged from 91% to 100%. The accuracy for MRI ranged from 48% to 100%.

Though there has been no weighting to the following summary values the overall sensitivity and PPV for MRI from the 9 studies as calculated from a summary 2x2 table is presented in Table 7.6.

Table 7.6: MRI per patient summary values and 2x2 table values

SENSITIVITY		80%		336 / 336 + 86
PPV		96%		336 / 336 +13
ACCURACY		91%		336+142 / 577
Total TP=336	Total FP=13	Total FN=86	Total TN=142	Total =577

TP: total positive results; FP: false positive results; FN: false negative results; TN: total negative results; Total: total number of patients investigated by MRI

PETCT data

The sensitivity for PETCT ranged from 91% to 100%. Specificity ranged from 60% to 100%. In a number of studies specificity estimates are not possible as there were no benign lesions identified at all in the population. The PPV ranged from 93% to 100%. Accuracy ranged from 91%-100%.

Though there has been no weighting to the following summary values the overall sensitivity and PPV for PETCT from the 6 studies as calculated from a summary 2x2 table is presented in Table 7.7.

Table 7.7: PETCT per patient summary values and 2x2 table values

SENSITIVITY		94%		273 /273+19
PPV		94%		273 / 273+19
ACCURACY		94%		273+153/453
Total TP=273	Total FP=8	Total FN=19	Total TN=153	Total = 453

TP: total positive results; FP: false positive results; FN: false negative results; TN: total negative results; Total: total number of patients investigated by PETCT

Per lesion analysis

7 studies reported CT data per lesion, 12 studies reported MRI data per lesion, 6 studies reported PETCT data per lesion.

CT data

The sensitivity of CT ranged from 67% to 97%. The PPV for CT ranged from 63%-100%. Specificity for CT ranged from 0 to 67%. In a number of studies specificity estimates are not possible as there were no benign lesions identified at all in the population. This finding might be expected in a population that is so highly selected for suspicion of malignancy. The accuracy for CT investigation ranged from 64% to 84%.

Though there has been no weighting to the following summary values the overall sensitivity and PPV for CT from the 7 studies as calculated from a summary 2x2 table is presented in Table 7.8.

Table 7.8: CT per lesion summary values and 2x2 table values

SENSITIVITY		74%		704 / 956
PPV		90%		704 / 792
ACCURACY		78%		704+114 / 1048
Total TP=704	Total FP=78	Total FN=252	Total TN=114	Total = 1048

TP: total positive results; FP: false positive results; FN: false negative results; TN: total negative results; Total: total number of lesions identified by CT

MRI data

The sensitivity of MRI ranged from 81% to 100%. Specificity ranged from 59% to 100%. In a number of studies specificity estimates are not possible as there were no benign lesions identified at all in the population. PPV ranged from 81% to 100%. The accuracy for MRI ranged from 71% to 100%.

Though there has been no weighting to the following summary values the overall sensitivity and PPV for MRI from the 12 studies as calculated from a summary 2x2 table is presented in Table 7.9.

Table 7.9: MRI per lesion summary values and 2x2 table values

SENSITIVITY	88%	1139 / 158		
PPV	96%	704 / 792		
ACCURACY	87%	1139+229 / 1571		
Total TP=1139	Total FP=45	Total FN=158	Total TN=229	Total = 1571

TP: total positive results; FP: false positive results; FN: false negative results; TN: total negative results; Total: total number of lesions identified by MRI

PETCT data

The sensitivity for PETCT ranged from 61% to 100%. Specificity ranged from 60% to 100%. In a number of studies specificity estimates are not possible as there were no benign lesions identified at all in the population. The PPV ranged from 94% to 100%. Accuracy ranged from 61%-100%.

Though there has been no weighting to the following summary values the overall sensitivity and PPV for PETCT from the 6 studies as calculated from a summary 2x2 table is presented in Table 7.10.

Table 7.10: PETCT per lesion summary values and 2x2 table values

SENSITIVITY	79%	410 / 522		
PPV	99%	410 / 415		
ACCURACY	97%	410+96 / 523		
Total TP=410	Total FP=5	Total FN=112	Total TN=96	Total = 523

TP: total positive results; FP: false positive results; FN: false negative results; TN: total negative results; Total: total number of lesions identified by PETCT

Summary sensitivity and specificity for all imaging modalities both for per patient and per lesion analyses are plotted and presented in Figures 7.2 and 7.3.

Figure 7.2: Per patient forest plot for colorectal liver metastases meta-analysis

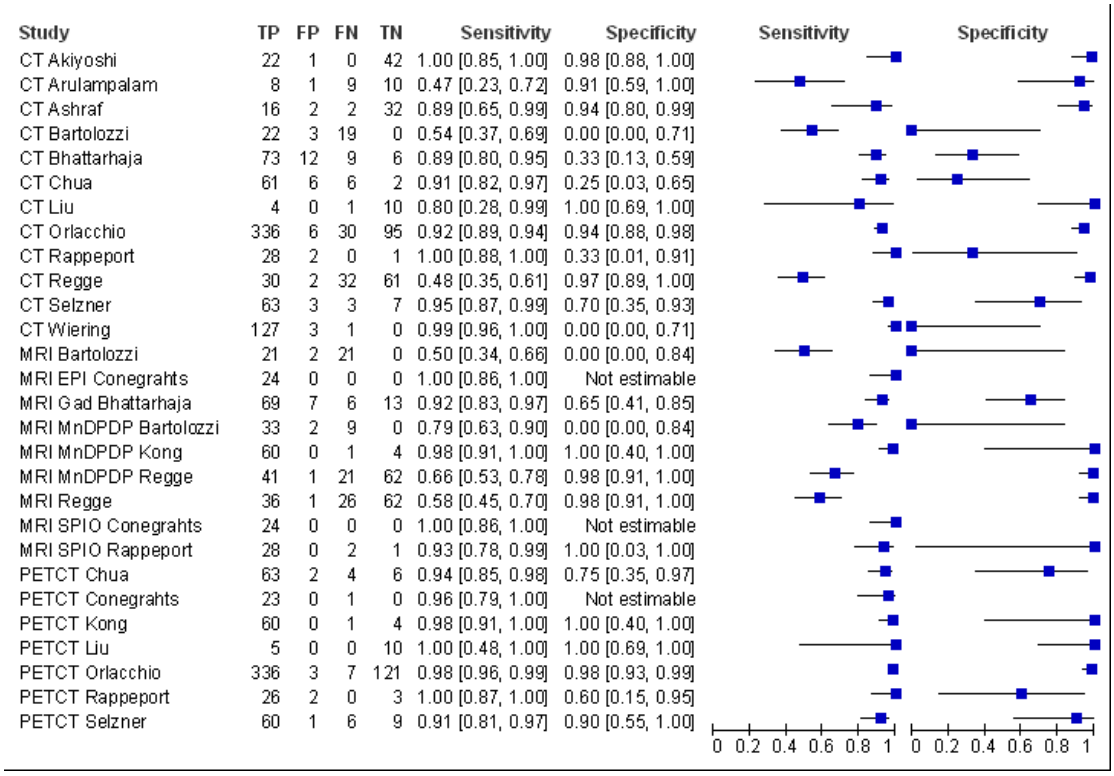
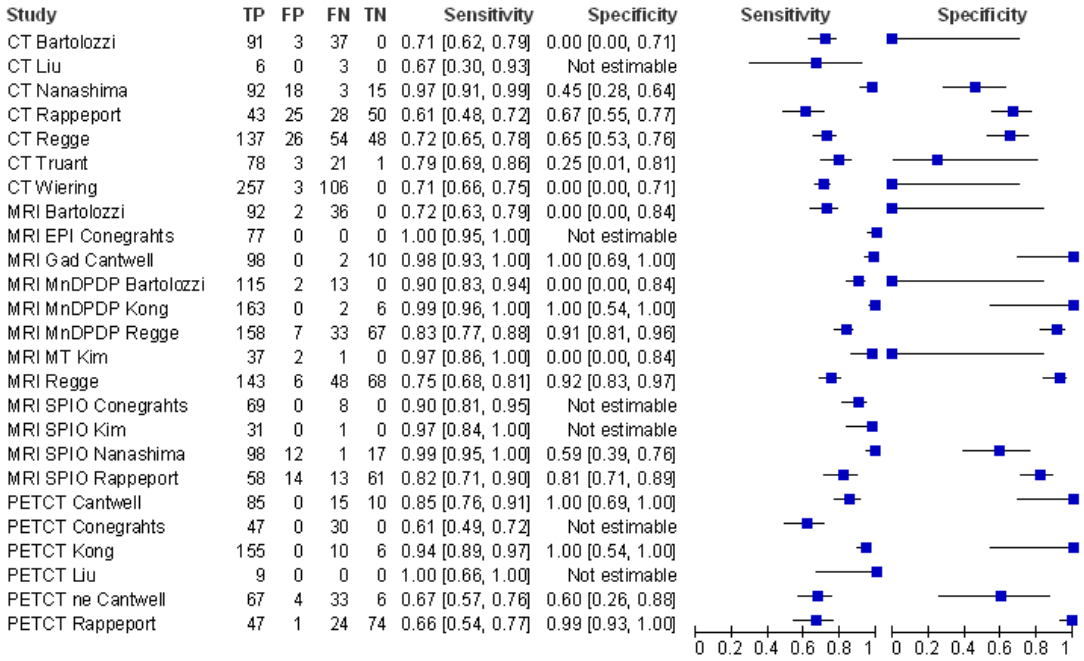


Figure 7.3: Per lesion forest plot for colorectal liver metastases meta-analysis



In many of the studies shown above specificity is not estimable. This is because these studies found no negative (benign) lesions. This is a plausible finding from meta-analysis of highly selected populations with patients already having suspicion of liver metastases on a previous scan.

7.5 Discussion

The results show that in a per patient analysis PETCT has consistently higher sensitivity in all the studies compared to MRI and CT. When pooling the data the summary sensitivity and accuracy for PETCT is 94% for both and is higher than MRI (80%,91%) and CT (87% for both).

In the per lesion analysis it is MRI that shows higher sensitivity compared with CT and PETCT. The pooled data shows MRI as having a combined sensitivity of 88% and accuracy of 87%, CT a sensitivity of 74% and accuracy of 78% and PETCT a sensitivity of 79% and accuracy of 97%.

Before any conclusions can be drawn from these results there are a number of methodological issues that need to be highlighted.

The higher diagnostic sensitivity of MRI in the per lesion analysis is potentially biased due to the data being clustered.

The parallel analysis of the data ‘per patient’ and ‘per lesion’ is advocated for clustered data, otherwise also known as longitudinal or correlated. There is a tendency of the measurements within a cluster to respond in a similar way, which introduces statistical complexity.[205]

When using a digital survey technique, such as PETCT, CT, or MRI, patients may present with multiple sites of tumour in their bodies. In sicker patients the presence of a lesion in a certain site increases the chances of observing a lesion in another site (positive correlation). [205]

Radiological experiments using the above mentioned modalities involve the collection of multiple observations for each patient studied. For a variety of reasons, different observations from the same patient are more likely to be correlated, while observations from different subjects can be considered statistically independent.[205]

Conventional statistical techniques are based on the assumption that all observations are independent of each other. Any conclusions made for correlated observations within a cluster based on the above assumption will therefore not be valid. The extent of the problem depends heavily on the magnitude of the correlation as well as on the number of observations within the cluster.[205]

One way of handling the problem of clustering is to provide a summary measure for each subject, i.e. the proportion of positive sites per patient (per patient analysis) and base the analysis solely on this measure. This implies that the subject, and not the measurements within the subject, is taken as the unit of analysis.[205]

Reduction of the data to one observation per subject is convenient because independence among units of analysis is achieved and standard analysis methods are thus applicable. There is loss of information however owing to the aggregation of data within each subject.[205]

When the correlation between the observations in the same patient is positive, ignoring the correlation might result in an erroneous conclusion.[205]

Another methodological issue that has an effect on these results is the heterogeneity of the studies.

Systematic reviews bring together studies that are diverse both clinically and methodologically. Therefore heterogeneity in the results is to be expected.

Heterogeneity is likely to arise through a variety of causes including the population characteristics, diversity in doses, length of follow-up, study quality. What matters is the extent to which this heterogeneity affects the conclusions of the meta-analysis.[206]

Unless we know how consistent the results are, we cannot determine the extent to which the findings of the meta-analysis can be generalised.[206]

The challenge is to consider all the potential sources of heterogeneity for a planned systematic review and decide at the outset whether its effect on the overall outcome is such that the source of the heterogeneity needs to become an exclusion criterion.

If the decision is made to pool the studies despite the heterogeneity then this issue can potentially be addressed by carrying out subgroup analyses post meta-analysis and thus testing in this way whether the results are any different having removed the potential source of bias.

For this review the sources of heterogeneity were considered. However, despite these it was still felt appropriate to include all studies and pool the data. A subgroup analysis would then be considered for each source of heterogeneity post meta-analysis.

Sources of heterogeneity that potentially could have affected the outcome of this review and could have under- or over-estimated the diagnostic accuracy of a modality are: the design of the studies, the quality of the studies, the inclusion of studies that did not exclude patients who had received neoadjuvant chemotherapy prior to their imaging, the size of the metastatic liver lesions, the number and experience of the radiologists reading the imaging results, the technological characteristics of the scanners, and population characteristics such as the presence of co-incident diabetes or other co-morbidities that could potentially give a particular group of patients an advantage or a disadvantage over another group of patients. These sources of heterogeneity will be discussed individually below.

Study Design

Diagnostic accuracy studies are less well understood and their design methodology is less robustly developed compared with therapeutic studies. The design across studies is known to be heterogeneous.[207-208]

The results from this review show that the great majority of studies included were prospective (n=15) and retrospective (n=5) case series and that there is a relative absence of randomised controlled trials (n=2) from this particular field of radiological oncology.

Although randomisation is usually the gold standard in the design of therapeutic trials for the investigation of diagnostic questions, it is often not straightforward. Randomisation is more appropriate for the investigation of the secondary outcomes of survival and change in patient management decisions. In fact one of the two randomised controlled studies that the search identified did in fact have these parameters as the primary outcome.[199]

For this review question in particular most studies had all patients and all potential lesions assessed by all index modalities under investigation so that their diagnostic abilities could be directly compared and they did not randomise any patient groups to one type of imaging modality.

Prospective rather than retrospective data collection is preferred in trial design because defining reference standard criteria and setting time periods allowed between tests and references standards is more precise and set in advance for the entire cohort but this is not always possible within clinical practice and so the type of data collection among studies is mixed.

There are also ethical limitations to diagnostic study design. For example giving every patient the same reference standard of histological examination of the resected specimen would be unethical if the index test has shown that a particular patient is inoperable or indeed has a benign lesion and therefore resection is contraindicated.

Study size also varies. The smallest study has 15 patients [194] and the largest 467 patients. [196]

Study quality

The quality of individual studies as assessed by the QUADAS tool also varied. Quality assessment is an integral part of any systematic review. If the results of the individual studies are biased and these are synthesised without any consideration of quality then the review will also be biased.[181]

This is particularly important for reviews of diagnostic accuracy studies as the methods for systematic data pooling and meta-analysis with the traditional application of weight to the different studies are still under development for diagnostic accuracy studies and this adds further potential source for bias.

One of the main findings of the quality assessment with the QUADAS tool was that of poor reporting of items in the individual studies. It is a recognised problem particularly in the field of diagnostic accuracy studies.

One of the items addressed by QUADAS is the time interval between index tests and reference standards. This time period potentially means that the disease may have progressed and misclassification of lesions may occur (disease progression bias). The length of period that may cause such bias is different between conditions. For cancer up to one month between tests and between 3-6 months for follow-up of a lesion is commonly accepted. 55% of studies failed to report the time elapsed between index and reference tests.

A number of items on the QUADAS checklist address the issue of ‘blinding’. Lack of ‘blinding’ of investigators participating in a trial is an established source of potential bias for all types of studies. Test review bias occurs when persons interpreting the test under investigation have knowledge of the gold standard / reference test result. Diagnostic review bias occurs when interpretation of the gold standard test is made with knowledge of the test under investigation. Clinical review bias occurs when there is availability of other relevant clinical information (e.g. symptoms, co-morbidities), which may also affect estimates of test performance.[209-210]

82% of studies did not report on whether the participating investigators interpreted the reference test without knowing the results of the index test and 41% of studies failed to report whether the investigators interpreted the index test without knowing the results of the reference standard.

The final items on the QUADAS checklist record whether studies report their withdrawals and indeterminate results. Authors may deal with these cases by considering them as positives, negatives or by excluding them from the analysis.

Each of these methods could potentially alter the reported test performance. Considering these cases positive increases sensitivity, considering them negative increases specificity. In any case, a high number of indeterminate results diminishes the value of a test. This review found that 75% of studies did not report their withdrawals and 85 % of studies failed to report indeterminate results.

Overall, the results of the quality assessment show that the majority of studies are of poor quality mainly due to poor reporting and to weaknesses in their design. This is in agreement with findings of previous reviews and unfortunately weakens the potential impact of the results.

A subgroup analysis of the better versus the poorer studies was not performed, as the poor quality was so extensive.

Neoadjuvant chemotherapy

It has been suggested that the use of chemotherapy up to a month prior to a PET-based examination may reduced the sensitivity of PET technology in detecting tumours[211], although others have challenged this shortcoming[212].

Patients may have had chemotherapy prior to their imaging either to downstage their primary tumour or as post-operative treatment for their

primary tumour. This may have an effect on the liver lesions if they are metastatic.

Lesions that are responding to chemotherapy treatment may not appear as well defined on PET-based scanning. The metabolism of the lesion is changed and this results in lesser or no appearance on the PET or PETCT scan. This could lead to higher number of false negatives for these modalities. Chemotherapy does not affect the identification of malignant pathology by CT or MRI.

Of the 22 studies included in this review 4 studies comment on chemotherapy-exposed patients [189,197,198,201] and of those only 2 actually give diagnostic data.

In the study by Chua et al [189] no data is provided for chemotherapy exposed and naive patients but the group comment that chemotherapy did not impact on the diagnostic accuracy of PETCT (p=0.178).

In the study by Regge et al [198] 19 patients out of 125 that were included in the study had neo-adjuvant chemotherapy but no details are given with regard to the timing of the chemotherapy and no separate data is given for this subroup of patients.

The Rappeport study [197] report that 4 out of the 35 patients had prior chemotherapy. The 4 patients had 14 liver metastases (7 more than 1cm). For liver metastases larger than 1cm, PET had a sensitivity of 43% (3/7) compared to 76% (34/45) in patients without recent chemotherapy. For liver metastases up to 1cm the result was 0% (0/7) versus 8% (1/12). Even when patients with recent chemotherapy were excluded from the analyses, CT and SPIO-MRI were significantly more sensitive than PET (p=0.001) There is no mention about the chemotherapy effect on PETCT.

In the study by Selzner et al [201] 18 patients received chemotherapy within a month prior to PETCT. A comparison of the false negatives revealed that FDG

uptake was absent in the liver metastases in 5 patients (28%) in the group with recent chemotherapy (sensitivity 72%) and in 1 in 58 of patients (5%) in the group without recent chemotherapy (sensitivity 98%, p=0.14 Fisher's exact test).

Overall the data available is very limited and therefore subgroup analysis looking at the association of neo-adjuvant chemotherapy and the diagnostic accuracy of PETCT was not carried out for this review.

Varied size of liver lesions

There is evidence from previous PET studies that PET technology is not as efficient at picking up micrometastases as it lacks the anatomical definition of CT and MRI.[213]

A subgroup analysis attempting to distinguish the diagnostic ability of the modalities in association with lesion size would be an appropriate way to handle this source of bias.

Of the 22 studies included in this review 7 comment on diagnostic accuracy in association with lesion size (186,187,198,203,204,193,197). Lesions are grouped into small lesions of <1cm in diameter, intermediate size lesions of 1-2cm, or larger lesions of >2cm.

In the study by Bartolozzi et al [186], in the lesion by lesion analysis for lesions ≤1cm the difference in sensitivity among spiral CT (38%), unenhanced MRI (51%), and MnDPDP-enhanced MRI (83%) was even more manifest than in the overall comparison of diagnostic sensitivity. All lesions undetected by MnDPDP-enhanced MRI and discovered by the reference test did not exceed 1cm in diameter.

In the Regge study [198] the reference standard detected 191 lesions, 35.1% of which were ≤10 mm in size; per-lesion sensitivity was 71.7, 74.9 and 82.7% for CT, unenhanced MRI and MnDPDP-enhanced MRI, respectively. Although

the sensitivity of MnDPDP-enhanced MRI for ≤10 mm lesions was higher than both CT and unenhanced MRI (67.7 versus 47.7% versus 53.8%, respectively), multivariate analysis showed that lesion size was not significantly associated with such differences.

In the study by Vidiri et al [203] of the 13 lesions that were ≤1cm CT identified 4, unenhanced MRI identified 2, and contrast-enhanced MRI 9. The group did not perform statistical analysis on these results. For both the small lesions group, and the larger lesions group the performance of the modalities was very similar with CT and MRI identifying 10 lesions of 1-2cm, enhanced MRI identifying 12, and all modalities identifying 20 lesions of the 21 lesions that were larger than 2cm.

In the study by Wiering et al [204] the comparison is between CT and PET with only the CT data relevant to this review. CT identified 10 of the 63 lesions that were ≤1cm (16%), 123 of the 172 lesions that were 1-2cm (71%), and 124 of the 128 lesions that were greater than 2cm. Both CT and PET in this study missed the majority of lesions smaller than 2cm, both missed 25% of lesions 1-2cm in size, and detected equally satisfactorily the lesions larger than 2cm.

In the study by Bhattarajha et al [187] the sensitivities of CT and contrast enhanced MRI for lesions ≤1cm are 52% and 57% respectively and for lesions >1cm the sensitivities are 77.4% and 86.3%. No formal subgroup analysis or conclusion is presented.

The study by Kong et al [193] is one of two only studies where there is a direct comparison of PETCT and MnDPDP-enhanced MRI. In this study it is reported that MRI correctly identifies more lesions in 8 scans compared with PETCT. The lesions not detected by PETCT were all ≤1cm apart from one that was 1.5cm. PETCT correctly identified more metastatic lesions than MRI in one case and correctly identified as a metastasis one equivocal MRI case.

Based on these descriptive results and with no reported statistical analysis the group conclude that MRI is superior to PETCT for the identification of small liver metastases and should remain therefore a prerequisite for surgical planning in patients with liver metastases.

The Rappeport study [197] is the second of the two studies that directly compare MRI to PETCT (as well as to PET and CT). Only one of 19 metastases that were ≤1cm in size was correctly identified by PET, 5 were identified by PETCT, whereas CT detected 13, and SPIO-MRI 10. Of the 52 metastases that were >1cm PETCT identified 42, PET 37, and SPIO-MRI 48. Overall, all modalities were more sensitive in detecting liver metastases that were >1cm compared with ≤1cm. No formal statistical analysis is made of these results. Overall the group conclude that PETCT equalled MRI imaging in accuracy for liver metastasis detection but made no further specific conclusion about association to liver lesion size.

As is evident from the presentation of the data from the 7 individual studies a conclusion that is repeated across the studies is that contrast-enhanced MRI is better than non-enhanced MRI and CT at identifying micrometastases. The subgroup analysis that would have been of interest comparing PETCT to contrast enhanced MRI was not carried out for this review as only two studies actually gave PETCT data specific for the identification of micrometastases. Both these studies, which individually are relatively small (Kong n=65 [193], Rappeport n=35 [197]) report contrast enhanced MRI having better diagnostic ability at identifying micrometastases compared with PETCT but no statistical analysis is done to confirm the significance of this result.

The number and experience of the readers

The diagnosis formulation is based on different radiologists across all the studies reading the images. They have different levels of experience and different abilities. The diagnostic ability of each modality is only as good as the diagnostic ability of the reader of the images produced.

A radiology consultant of 15 years of experience with an interest in colorectal radiology who sub-specialises in MRI only radiology and works at the national specialist centre may have a different diagnostic ability to a consultant of <5 years in general radiology who works in a district general hospital. Some studies may counteract this effect by having the images read by more than one radiologist, and where the two disagree, then base the final diagnosis on their consensus opinion.

Out of the 22 studies in this review only 9 studies comment on the number of radiologists that read the index tests, only 1 study mentioned the experience of the radiologist, and only 2 studies provided kappa statistics to quantify the level of agreement between the two or more radiologists. Subgroup analysis based on this parameter therefore was not carried out for this review.

Technological aspects of the index test
The imaging modalities are heterogeneous in their technologies both in the principle of their diagnostic method and in how they are developed over the years. Slice thickness, amount and type of contrast used, strength of magnetic field applied are some of the characteristics that have changed over time.

The different scanners, as they develop technologically and as different centres use slightly different scanners, contribute to the heterogeneity of the included studies. All the aspects of the technology of each modality mentioned below can potentially affect the outcome.

For CT the type of scanner (helical, multi-section helical), the section thickness (<5mm or >5mm), the amount of contrast agent (<45g iodine or >45g iodine), the number of phases (1=portal or 2=arterial and portal) are all characteristics that contribute to heterogeneity.

For MRI the magnetic field strength, the type of contrast agent used (non-specific or liver specific, non-enhanced, gadolinium enhanced, SPIO-enhanced), the sequences, the type of coil used (body coil or phased array

coil), the section thickness are all characteristics that can also contribute to heterogeneity.

For PET the system type (dedicated full ring or other), the amount of tracer used, the type of analysis (qualitative or quantitative), the data acquisition characteristics (timing of scanning and time of scanning per table position) are characteristics that contribute to heterogeneity.

For this review the data was too varied to make subgroup analysis worthwhile on this topic.

Population characteristics

Differences between populations in demographic and clinical features may produce measures of diagnostic accuracy that vary considerably; this is known as spectrum bias. Reported estimates of diagnostic test accuracy may have limited clinical applicability (‘generalisability’) if the spectrum of participants tested is not representative of the patients on whom the test will be used in practice.[180]

For this review all studies were consistent in including a representative group of patients. All studies included looked at data relating to patients both female and male, with a confirmed diagnosis of colorectal cancer only, and either confirmed lesions or lesions suspicious of liver metastases. Studies which reported on diagnostic accuracy of the modalities of interest but did not distinguish between liver metastases from colorectal cancer and other cancers were excluded. The age of the population, their co-morbidities, the referral patterns, the diagnostic settings are also similar between the studies and the population of interest. Therefore heterogeneity in the sources of the study population can be accepted as minimal.

However, a population characteristic that may potentially introduce heterogeneity is if patients included were diabetic. PET works on the principle of altered metabolism of glucose in patients with cancer. Most studies include

patients who have well-controlled diabetes but this is theoretically still a source of spectrum bias.

There are other sources of bias which add to the methodological issues relevant to these results.

Selection bias

Selection bias refers to the distortion of a statistical analysis resulting from the method of collecting samples. If the selection bias is not taken into account then certain of the conclusions drawn may be incorrect.[214]

The start date for the search of the medical literature that was carried out for the purpose of this systematic review was set as 1995. This was on the advice of experts in radiology who suggested that to include CT, MRI and PET technologies prior to this date would not be appropriate. Though PETCT specifically is a much more recent technology than CT or PET, the start date for the search was not brought forward based alone on the availability of this imaging method but instead the *a priori* decision was taken to include only in the meta-analysis studies which reported comparisons using PETCT and not those using the older PET scanners.

Although two previous meta-analyses had pooled data with a search end date of 2004 the search for this review was not started post 2004 because this would not have been a strong enough guarantee of the absence of PETCT studies prior to 2004. All these *a priori* decisions potentially include selection bias but in the context of advancing technology they were deemed appropriate for this review.

Differential verification bias

All the studies are subject to differential verification bias. This is because some of the index test results are verified by a different reference standard. This is a problem as the reference standards differ in their definition of liver metastasis. Histopathology has a far more precise definition of malignancy and is the gold standard compared with repeating the imaging and defining

malignancy based on change in size of a lesion. However for this particular area of oncology and surgery it would be unethical to submit a patient unnecessarily to the risks of surgery or indeed to the risks of a biopsy in the liver for benign or inoperable malignant lesions. Therefore it is acceptable that the studies were set up in this way even though the results are subject to verification bias.

Incorporation bias

This occurs when results of the test under study are actually used to make the final diagnosis.[210] The QUADAS results show that all studies included had a different reference test to the index test used so are not subject to incorporation bias.

Publication bias

As the data used in studies of test accuracy are often collected as part of routine clinical practice (and in the past have tended not to require formal registration) it has been argued that test accuracy studies are more easily conducted and abandoned than RCTs. They may therefore be particularly susceptible to publication bias.[210] It has been demonstrated however, that the unique features of the test accuracy study make the application of tests of funnel plot asymmetry potentially misleading.[210] It should be noted that the power of all statistical tests for funnel plot asymmetry decreases with increasing heterogeneity. Given the limitations of current knowledge, to ignore the possibility of publication bias would seem unwise; however, its assessment in reviews of test accuracy is complex [210] and it has not been undertaken for this review.

Secondary outcomes

A further methodological concern regards the reporting of secondary outcomes. 6 studies in this review report on change in management [183,184,189,194,199,201]. 3 studies [183,184,199] include PET and the other 3 studies [189,194,201] include PETCT as the index test. All of the

studies agree that it is the PET-based modality that is responsible for the change in management.

However, all of the studies incorporate the extra-hepatic diagnostic element of these modalities in the results and in fact report that it is the detection of extra hepatic disease by these modalities which accounts for 25-40 %[189,194] in the change in management.

In a review such as this which addresses a question on hepatic metastases specifically it is not appropriate to include and to analyse results that are the reflection of more complex data collection. It is sometimes possible to separate multiple sets of data from a study if the authors' presentation allows for this. For example, this is often the case when studies present diagnostic data of a particular imaging modality for both lower GI cancer patients and patients with upper GI cancers.

In this review and with regard to the secondary outcome of change in management it has not been possible to separate the hepatic data from the extra-hepatic data in any of the studies and indeed it would not be clinically appropriate to do so.

Patient management is most often multifaceted. Attempting to decipher which scanner is best for detecting whether hepatic metastases are within the limits of what is resectable provides an answer to an essential part of the decision-making pathway for patients with metastatic colorectal cancer. But when considering the patient's overall management this step is linked to the parallel exclusion of the presence of extra-hepatic metastases, particularly pulmonary and peritoneal, as this is a potential contraindication to surgery.

The secondary outcomes of survival and change in management cannot be reported on in association with hepatic metastases in isolation in this review. This is in agreement with previous meta-analyses which have shown that in their great majority the design of diagnostic studies in this field provide results which cannot be analysed with regard to one type of metastasis in isolation.

This is important when considering the overall diagnostic impact of a specific modality.

As a consequence the feasibility of conducting any comprehensive analysis of cost-effectiveness to accompany this review is also therefore limited as it would need to take into account not only the diagnostic accuracy of the imaging modality in detecting hepatic metastases, but also the subsequent treatment decisions and patient outcomes.

As the search of the clinical literature revealed that most of the relevant studies identified do not report information on change in patient management in relation to the information obtained by the imaging test and as the decision to resect is based on a number of different considerations, there is insufficient information to model the link between the imaging results and the treatment decision.

In conclusion, from the evidence all modalities are comparably very good at detecting liver metastases with sensitivities above 75%. However, PETCT reported higher sensitivity than the other modalities in the per patient analysis. Contrast-enhanced MRI appears to have superiority in detecting micro-metastases as indicated by a small proportion of the studies.

However, these results are significantly affected by the overall poor quality of the pooled studies, which are subject to multiple sources of bias. This is mainly due to poor reporting of study design parameters and an overall variation in study design.

Therefore the available evidence is unclear as to whether MRI or PETCT should be used after a CT scan to confirm the patient with liver metastases suitable for surgery, and further research is necessary.

A prospective trial should be conducted to investigate the most clinically effective and cost-effective sequence in which to perform MRI and PETCT,

after an initial CT scan, to determine whether metastases are resectable in patients with colorectal cancer that has metastasised to the liver.

The diagnostic accuracy of a given modality in identifying either hepatic or extra hepatic metastases is an appropriate and desirable study outcome. However these diagnostic events are clinically linked when considering surgery for liver metastases. Therefore, in trying to identify which modality is best for staging patients for liver surgery it might not be the direct diagnostic accuracy of each modality that is important but rather the downstream effect it has on subsequent clinical decisions and on overall survival.

Therefore the outcomes of interest for future research should include reduction in inappropriate laparotomies, change in management decisions, and improvement in overall survival.

Subgroup data collection is necessary, particularly based on lesion size, if uncertainty for the need for multi-modality treatment (PETCT and MRI) is also to be resolved. The effect of neo-adjuvant chemotherapy and reader agreement should also be included in the data collection for subgroup analysis to be possible.

Some patients with colorectal cancer and liver metastases have a significant chance of cure with accurate and focused treatment. At the same time those patients whose disease is too advanced should not have the quality of their remaining life reduced by unecessary operations.

Therefore delineating the most successful pre-operative staging modality for patient selection for surgery is of utmost importance.

8.0 The availability and quality of the therapeutic data that make up the evidence to support the most effective method for the follow-up of patients who have been diagnosed and treated for primary CRC

8.1 Introduction

Valid guidelines create their evidence components from systematic reviews of all the relevant worldwide literature. Some recommendations may be derived from evidence of high validity and others from evidence that is much more liable to error.[215] Sources of evidence can range from small laboratory studies or case reports to well-designed large clinical studies.[216]

Checklists assessing important methodology parameters relevant to the quality of a systematic review [217] or a randomized controlled trial [218] for example have been traditionally used by some guideline-producing organizations such as NICE in order to assess systematically the internal validity of evidence being used.

The checklists, through a series of questions, aim to establish the quality of the internal validity of the study under review. That is, to make sure that it has been carried out carefully, and that the outcomes are likely to be attributable to the intervention being investigated. Each question covers an aspect of methodology that research has shown makes a significant difference to the conclusions of a study.

Evidence grading schemes have been used for over 25 years. Since the 1970's a growing number of organizations have employed various systems to grade the quality (level) of evidence and the strength of guideline recommendations.[216]

Some grading systems are based on study design alone without explicit consideration of other important factors in determining the quality of evidence. Some systems are excessively complex.[216] A commonly used system is presented in table 8.1.

Table 8.1: The University of Aberdeen system for the grading of evidence [219]:

	Grading of evidence
Ia	Evidence obtained from meta-analysis of randomized controlled trials
Ib	Evidence obtained from at least one randomized controlled trial
IIa	Evidence obtained from at least one well-designed controlled study without randomization
IIb	Evidence obtained from at least one other type of well-designed quasi-experimental study
III	Evidence obtained from a well-designed non-experimental descriptive study, such as comparative studies, correlation studies and case studies
IV	Evidence obtained from expert committee reports or opinions or clinical experiences of respected authorities
	Grading of recommendations
A	Evidence categories Ia and Ib
B	Evidence categories IIa, IIb, III
C	Evidence category IV

Unfortunately, different organizations use different systems to grade evidence and recommendations. The same evidence and recommendation could be graded as “II-2”, “B”, “C+”, “1”, or “strong evidence, strongly recommended” depending on which system is used. This is confusing and impedes effective communication.[216]

Additionally with this system of evidence grading any data that has been derived from a systematic review of randomized controlled trials for example could potentially be given the top grade – level I – of evidence based on this parameter alone.

However recent methodological research has lead to a more extensive method of quality assessment and evidence grading. GRADE (Grading of Recommendations Assessment Development and Evaluation)[220] is a new system of grading evidence.

The GRADE Working Group in 2000 developed a common, sensible and transparent approach to grading quality of evidence and strength of recommendations. Many international organizations have provided input into the development of the approach and have started using it.[220].

Before assessing the quality of the evidence with GRADE, systematic reviewers and guideline developers identify all outcomes of a study or review, including benefits, harms, and costs. Reviewers will then assess the quality of evidence for each important outcome.[221] The GRADE-pro software is used to compare studies both in terms of internal validity and also of outcomes presented.[216]

With this method a systematic review of randomised controlled trials may have been conducted with high quality but when in addition to its internal validity assessment all of the outcomes within the review are assessed it may be that despite a high quality systematic review some of the outcomes are given a low grading.[216]

Table 8.2 presents the way in which evidence from any given review of the literature is classified by GRADE.

Table 8.2: The GRADE system of classification of the quality of evidence.[216]

High quality	Further research is very unlikely to change our confidence in the estimate of effect
Moderate quality	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low quality	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low quality	Any estimate of effect is very uncertain

The approach to rating the quality of evidence with GRADE begins with the study design and then addresses five possible reasons to rate down the quality of evidence and three possible reasons to rate up the quality.[221]

The five reasons to rate down the quality of evidence are: risk of bias, imprecision (relating to how narrow or wide the confidence intervals are or how sparse the data is), inconsistency (or heterogeneity), indirectness (or applicability), and publication or reporting bias.[221]

The three reasons to rate up the quality are: if the magnitude of the treatment effect is very large; if there is evidence of a dose-response relation; or if all plausible biases would decrease the magnitude of an apparent treatment effect (if, for instance, only sicker patients receive an experimental intervention or exposure, yet they still fare better, it is likely that the actual intervention or exposure effect is even larger than the data suggest.[221]

GRADE identified these reasons because they address nearly all issues that bear on the quality of evidence. These categories were arrived at through a case-based process by members of GRADE, who identified a broad range of issues and factors related to the assessment of the quality of studies. All potential factors were considered, and through an iterative process of discussion and review, concerns were scrutinized and solutions narrowed by consensus to these five categories.[221]

For recommendations the GRADE system offers two grades: “strong” or “weak”. When the desirable effects of an intervention clearly outweigh the undesirable effects, or clearly do not, guideline panels offer strong recommendations. On the other hand, when the trade-offs are less certain (either because of low quality evidence or because evidence suggests that desirable and undesirable effects are closely balanced), weak recommendations become mandatory.[216]

The GRADE system is today used widely. The World Health Organization, the American College of Physicians, the American Thoracic Society, UpToDate (an electronic resource widely used in North America - http://www.uptodate.com), the Cochrane Collaboration and NICE are among the organisations which have adopted

GRADE. NICE specifically recommends the use of GRADE in the assessment of therapeutic data / evidence.

One such topic of therapeutic evidence included in the NICE colorectal cancer guideline is that of the most appropriate follow-up strategies for CRC. After definitive treatment for CRC is completed, the attention of clinicians turns to follow-up strategies designed to detect tumour recurrence at a stage when further curative procedures can be used. Follow-up strategies have also been developed in order to detect curable metachronous (i.e. occurring at different times) tumours. There is overlap between these two strategies.[222]

The opportunity cost of the resources involved is considerable.[223,224] Clinicians justify follow-up by claiming that recurrences are being detected earlier than would otherwise occur and that patient outcomes are improved as a result.[223] Routine follow-up has the potential to create psychological harm for patients and any such disadvantages need to be outweighed by improved clinical outcomes that matter to patients.[225]

Whether systematic follow-up can alter long-term clinical outcomes for CRC remains controversial. Whilst some commentators have concluded that follow-up is worthwhile [224], others have questioned its effectiveness.[223,225]

It has not yet been shown whether assessment of the body of evidence for a particular topic of a guideline using the new method of evidence grading does in fact make a difference to the results presented to a GDG and thus help to identify better the strengths and weaknesses of the given evidence.

8.2 Aim

To research the availability and quality of the therapeutic data that make up the evidence to support the most effective method for the follow-up of patients who have been diagnosed and treated for primary CRC.

8.3 Methods

A systematic review was performed of the most recently published studies on the topic of the effectiveness of follow-up strategies in CRC patients treated with curative intent.

The systematic review was carried out by one reviewer (the author) who selected the relevant titles and abstracts from the results of the database search, decided which studies met the inclusion criteria, and graded their methodological quality.

The body of evidence was assessed and classified using both the traditional NICE methodology checklists (Appendices 4 and 5) and GRADE. For the assessment with GRADE the relevant data was extracted and entered into GRADE pro software.

8.3.1 The PICO

The PICO for a therapeutic review follows the traditional framework of population – intervention – comparison – outcomes as detailed in table 8.3 below.

Table 8.3: The PICO for the systematic review on the follow-up of CRC

Population	Intervention	Comparison	Outcomes
Asymptomatic patients who have undergone treatment with curative intent for CRC, including patients treated for metastatic cancer	Intensive packages including: Clinical examination CEA tumour marker tests Imaging Colonoscopy Timing / duration	Do nothing Less intensive packages	Survival Recurrence Quality of life Metachronous primaries Late effects

Population

Males and females of any age with histologically-proven adenocarcinoma of colon or rectum, staged as T1,2,3,4 and treated with curative intent.

The decision was taken to include also patients who were treated for metastatic colorectal cancer but with curative intent. This is because advances in the treatment of metastatic disease mean that more patients have the chance of cure than ever before.

Evidence shows that neoadjuvant combinational chemotherapy may downstage up to 22% of unresectable liver metastases, hence increasing the proportion suitable for liver resection with curative intent.[227]

It has also been shown that early systemic chemotherapy for asymptomatic metastatic colorectal cancer improves the interval to symptomatic deterioration, compared with delaying chemotherapy until symptoms develop.[228]

Intervention

The liver is the most common site of metastases from colorectal cancer. A number of strategies have been proposed to detect liver metastases at an early stage in order to identify patients suitable for curative liver resection.

These include monitoring of blood tests (liver function, serum carcinoembryonic antigen (CEA)) and routine imaging of the liver [229,230] at various intervals during the first few years after initial curative resection.

Patients with resected colorectal cancer are also at risk for metachronous neoplasms in the colon and rectum (including second primary cancers) and of intra-luminal recurrent disease.

Patients with surgically resected Stage I, II and III cancers, and Stage IV cancer resected for cure (isolated hepatic or pulmonary metastasis) are candidates for

endoscopic surveillance, and the first preference is colonoscopy. The baseline is the ‘clean colon’ examination. There is variation in the literature on guidance for the optimal timing of this.

The duration to be recommended for intensive surveillance programmes is also unclear. The most important phase of follow-up is the first 2 - 3 years after primary tumour resection, as during this time the majority of recurrences will become apparent.[231-232]

The conventional ‘model’ was that follow-up stopped after 5 years and the patient was discharged as ‘cured’. The perspectives of this model are changing.

Specifically, in common with the principle of long-term survivorship, there is a continuing need to audit and quantify late effects from cancer treatment – for example, anorectal dysfunction from pelvic radiation.

In addition, it is now recognized that the incidence-rate of second primary colorectal cancers after the index case is constant over 15 years (with the possible exception of the first two years).[233]

Comparator

The variation in follow-up programmes used by clinicians is considerable.[233-7] Which of these is the optimal one remains unclear and provisional conclusions change with advancing technology.

The optimal setting for the follow-up of patients after curative treatment for colon and rectal cancer is also unclear. In many units patients attend a colorectal clinical led by the colorectal specialist supported by a colorectal cancer nurse specialist (CNS).

There is debate whether patients can be followed in the community using for example a telephone follow-up scheme or a GP-led surveillance regimen.

The studies included in this review compare different follow-up strategies. These included comparisons of follow-up versus no follow-up, follow-up strategies of varying intensity and follow-up in different healthcare settings.

Outcome

The primary outcome is overall survival. However there are a number of secondary outcomes that are equally important.

These include disease-specific survival, time to diagnosis of recurrence, incidence of surgery (with curative intent) for recurrence, interval between planned visits for assessment for recurrences, quality of life, harms, and costs of surveillance and investigations.

8.3.2 Search

Table 8.4 lists the databases that were searched for this PICO. The truncation symbols incorporated in the search are the same as those that have been presented previously relating to the search carried out in chapter 7 (Table 7.2).

The author closely observed the NCC-C information specialist create and perform the detailed search having received training in the creation of search strategies through the department of evidence based healthcare at Oxford university.

Members of the guideline development group for the NICE colorectal cancer guideline with professional experience in the area of colorectal cancer follow-up suggested that the literature search need not extend further back than 2005.

This was because there was a wealth of studies on the topic but these had been repeatedly reviewed and synthesised in recent years. Therefore the *a priori* decision was taken to perform a systematic review of the literature starting in 2005.

Table 8.4: List of databases searched for the systematic review on colorectal cancer follow-up with search dates

	Dates Covered	No. of references retrieved	Finish date of search
Medline	2005-2009	79	18/08/09
Premedline	2005-2009	18	19/08/09
Embase	2005-2009	68	19/08/09
Cochrane Library	2005-2009	76	17/08/09
Cinahl	2005-2009	5	19/08/09
BNI	2005-2009	9	19/08/09
Psychinfo	2005-2009	2	19/08/09
Web of Science (SCI & SSCI) and ISI Proceedings	2005-2009	14	19/08/09
Biomed Central	2005-2009	2	19/08/09
Total References retrieved (after de-duplication)		190	

The search strategy was as follows:

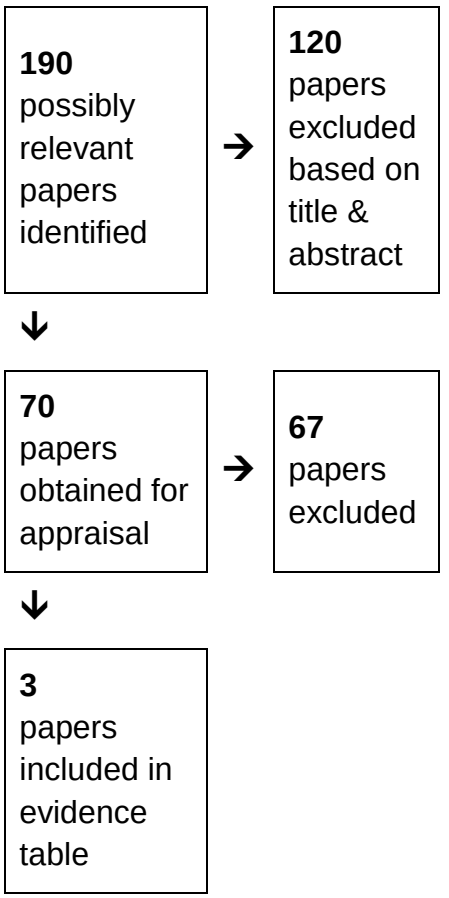
Colorectal Cancer And Follow UP

1. exp colorectal neoplasms/
2. ((colorect$ or colo rect$) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$)).tw.
3. ((colon or colonic) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$)).tw.
4. ((rectal$ or rectum$) adj3 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$)).tw.
5. 1 or 2 or 3 or 4
6. exp Follow-Up Studies/
7. (follow up$ or follow-up$).tw.
8. surveillance*.tw.
9. monitor*.tw.
10. 6 or 7 or 8 or 9
11. 10 and 5

8.3.3 Study selection

Study selection followed the steps displayed in the flow chart of Figure 8.1.

Figure 8.1: Flow chart showing the selection criteria for included evidence for the systematic review on colorectal follow-up



Reasons for exclusions:

- Opinion papers or narrative reviews that were not systematic were excluded (10).
- Papers not relevant to PICO were excluded (46).
- Reports of experimental tests for follow-up were excluded (8).
- Guidelines were excluded if the grading of evidence was unclear (3).

8.3.4 Critical appraisal of included studies

Methodology checklists

The NICE methodology checklists for systematic reviews (Appendix 4) and randomised controlled studies (Appendix 5) were used to assess internal validity in the traditional way.

The checklist on systematic reviews includes three sections. The first examines the internal validity of the study. It has five topics of investigation which focus on the study question, the description of the methodology, the description of the literature search, the assessment of study quality and the presence of heterogeneity. The second section assesses bias and its effect on the overall study result. The third section refers to the types of studies included in the review and how the overall conclusion answers the key question asked.

The answers are chosen from a list of six options (well covered, adequately addressed, poorly addressed, not addressed, not reported, not applicable).

The checklist on randomised controlled trials also includes three sections. The first section examines the internal validity of the studies. It has ten topics of investigation which focus on the study question, the randomisation, the concealment method, the 'blinding' methods, the similarity between the treatment and control groups at the start of the trial, the maintenance of the trial principles (that the treatment remains the only difference between the groups), the outcomes, the drop-out rate, the intention-to-treat analysis, and that results are comparable between different sites if the trial is of multi-centre type. The answers are chosen from the list of six options similar to the checklist for systematic reviews described above. The second section assesses bias and its effect on the overall study result. It also assesses applicability. The third section refers to the details of the study such as patient numbers, characteristics, comparators, follow-up period, outcomes, size of effect, funding.

GRADE

Table 8.5 summarises the GRADE approach to rating the quality of evidence, which begins with the study design (trials or observational studies) and then lists five possible reasons to rate down the quality of the evidence and three possible reasons to rate up the quality.

This is the basis on which the GRADE pro software carries out the calculation of the GRADE score and produced a final overall quality score for each study outcome.

Table 8.5: Rating of evidence by GRADE[221]

Study design	Initial quality of the body of evidence	Lower if	Higher if	Quality of a body of evidence
Randomised trials	High $\rightarrow$	Risk of bias	Large effect +1 large +2 very large	High (four plus ⊕⊕⊕⊕)
		Inconsistency	Dose response +1 evidence of a gradient	Moderate (three plus ⊕⊕⊕)
Observational studies	Low $\rightarrow$	Indirectness	All plausible residual confounding	Low (two plus ⊕⊕)
		Imprecision	+1 would reduce a demonstrated effect	Very Low (one plus ⊕)
		Publication bias	+1 would suggest a spurious effect if no effect was observed	

8.4 Results

The study selection resulted in 3 studies being included for data extraction. These were two systematic reviews of RCTs [222, 238] and one randomized controlled trial. [239]

8.4.1 Quality assessment by NICE methodology checklists

Quality Assessment of systematic reviews by NICE methodology checklists

Below are the summarised responses to the checklist questions both for systematic reviews as collected using the NICE methodology checklists (Appendix 4), followed by the evidence summary produced by the reviewing author without use of the GRADE pro software.

Question addressed
The two systematic reviews included both addressed a focused question that is directly relevant to the guideline PICO.

Applicability
Both systematic reviews refer to populations similar to that of the guideline PICO. That is asymptomatic patients with colorectal cancer who have been treated with curative intent. Patients may or may not have had adjuvant treatment. The populations are not identical as in the population of the studies no patients with metastatic disease were included although the PICO of this systematic review also aimed to address patients with metastatic cancer as long as their treatment was with curative intent. It was considered that this difference in the populations was not significant enough to affect the applicability of the study results.

Methodology
Both systematic reviews included clear and thorough descriptions of their methodology, which was appropriate. Only RCTs were included.

Literature searches of the two systematic reviews were appropriate and rigorous.

Study quality assessment

Both reviews assessed the quality of each individual study they included.

Heterogeneity

There is clinical heterogeneity of the follow-up regimen used by the different trials and this affects both systematic reviews equally. Some of the trials compared intensive follow-up with less intensive follow-up and some with no follow-up at all. In addition the intensity of the intensive follow-up regimen was variable and indeed the intensive regimen of one study was equal to the less intensive regimen of another study. In addition the trials span a considerable time period in which the surgery and oncological management of colorectal cancer has changed. In some studies patients were receiving adjuvant chemotherapy and radiotherapy and in others not. This adds to the clinical heterogeneity.

Overall Assessment

Both systematic reviews fulfilled all the criteria of the quality assessment checklists. None of the quality issues identified were thought likely to alter the conclusions of the reviews.

<u>Quality Assessment of the randomised controlled trial by NICE methodology checklists.</u>

Below are the summarised responses to the checklist questions for RCT as collected using the NICE methodology checklists (Appendix 5), followed by the evidence summary produced by the reviewing author without use of the GRADE pro software.

Question addressed

The RCT by Wang et. al. addressed a focused question that was directly relevant to the guideline PICO.

Recruitment / Applicability

The population of the study was similar to that of the PICO but excluded patients with metastatic CRC. There were clearly stated inclusion and exclusion criteria that were appropriate for the question being addressed.

Randomisation

All consecutive patients with newly diagnosed colorectal cancer who consented were randomised to the two trial groups. This was clearly explained and appropriate.

Allocation concealment

This was described clearly and was appropriate by means of sealed envelopes that contained cards allocating patients to one or the other arm of the study.

Maintenance

The status of the two study groups was comparable. The patients in both groups received similar management throughout the study period and the treatment they received was the only addition to their management.

The follow-up period was also clearly addressed in the study. The drop out rate was mentioned and within a normal range for a study of this nature. The analysis was on an intention-to-treat basis.

'Blinding'

In the study write up there was no reference to any 'blinding' having being undertaken.

Outcomes

All outcomes measured were relevant and measured in a standard, valid and reliable way. Complications were included in the outcomes.

Overall Assessment

This RCT fulfilled most criteria of the quality assessment checklist. It was not considered likely that the criterion which was not fulfilled ('blinding') would alter the conclusions of the study, although it potentially introduces bias.

Evidence Summary (without the use of the GRADE pro software)

Intensive versus less intensive follow-up:

There is significant overall survival benefit at 5 years with intensive follow-up.
The number of all recurrences detected is similar with both intensive and less intensive follow-up.
The number of curative procedures attempted for recurrence is significantly more with intensive follow-up.
There is no disease-specific survival benefit with intensive follow-up.
Significantly more asymptomatic recurrences were detected in the group which received intensive follow-up.
The time to recurrence is significantly less in patients receiving intensive follow-up.

Generally:

CEA blood test offers survival advantage.
Liver imaging offers survival benefit, CT of the liver improves survival, ultrasound scanning of the liver versus none offers survival advantage.
Colonoscopy leads to survival benefit versus no colonoscopy. Intensive colonoscopy versus less intensive colonoscopy does not offer survival advantage.
Clinic visits for follow-up versus no clinic visit offer survival benefit.

Complications:

1 study reported adverse events from follow-up relating to colonoscopy: 2 perforations and 2 GI bleeds from a total of 731 colonoscopies.[239]

Quality of Life (QoL):

1 study showed a small but significant increase in QoL with follow-up.[240]
1 study showed no difference in QoL in patients followed up in GP versus hospital setting.[241]

8.4.2 Quality Assessment by GRADE

The quality assessment by GRADE is presented in evidence tables produced by the GRADE pro software. These are included in Appendix 7.

There are 8 tables each presenting the assessment of evidence by GRADE for the 8 comparisons found in the studies identified by the literature search.

A question is presented per table as the formulated follow up comparison query. The table then presents the relevant associated outcomes for each question with data derived from the three studies included in this review.

The penultimate column for each table represents the overall GRADE quality score for that outcome.

The systematic review by Jeffery et al presented the results as odds ratios but the review by Tjandra et al presented the results in the form of p values.

For ease of comparison of results between the two systematic reviews where p values were reported in the Tjandra publication the odds ratio has been calculated and this is the result that has been entered in the GRADE software.

Calculations performed for this conversion are tabulated at the end of the GRADE evidence tables (Appendix 7).

These results did not have a numerical confidence interval reported and this is missing from the grade tables but the grading of precision and statistical significance has been made based on the confidence interval representation on the forest plots.

Evidence summary using the GRADE pro software.

Intensive v less intensive follow-up

There is moderate quality evidence of significant overall survival benefit at 5 years with intensive follow-up.[222, 238]

When looking at disease specific survival there is low quality evidence of survival benefit with intensive follow-up. The result is imprecise and not statistically significant.[222]

There is moderate quality evidence that the number of all recurrences detected is similar with both intensive and minimal follow-up.[222,238]

There is low quality evidence that significantly more asymptomatic recurrences were detected in the group that received intensive follow-up. The result is statistically significant though the total number of events in the pooled comparison is low and this can introduce imprecision to the result.[222,238]

The time to recurrence is significantly less with intensive follow-up but the evidence is of low quality mainly as the studies pooled are too heterogeneous.[222,238]

There is low quality evidence that the number of curative procedures attempted for recurrence is significantly greater with intensive follow-up.[222,238]

Clinic visits versus no visits or fewer clinic visits

Clinic visits versus no clinic visits showed a survival benefit in the clinic group but the result is imprecise, is not statistically significant, and the quality of the evidence is low.[222]

In the ‘more clinic visits’ versus ‘fewer clinic visits’ comparison a survival benefit was shown in the group attending more clinic visits both in terms of overall survival and

number of recurrences but again the result is imprecise and the quality of the studies low (recurrences) and very low (overall survival).[222]

More versus less frequent tests

There is evidence of moderate quality to support the contention that generally when more tests are done they do give a significant survival advantage over fewer tests done but the number of recurrences detected as a result is no different in the two groups.[222]

Carcinogenic Embryonic Antigen (CEA) (blood test)

When included in the follow-up protocol CEA estimation gives an overall survival advantage versus no CEA testing. However this result should be interpreted with caution.

The review by Jeffery et al reports this result as of no statistical significance. Only one randomised control trial [242] contributes to this result and overall the quality of the evidence is low. The RCT is small (107 patients) and the result is imprecise with a wide confidence interval.

The review by Tjandra et al reports a statistically significant result. Two RCTs have been pooled for this result [242,243] but the quality of the evidence is very low. There is serious imprecision. The confidence interval is only given as a diamond representation on the forest plot and not as a numerical value but the total number of events is low (less than 300).

The number of recurrences detected is not higher in the CEA group but this was not a significant result and the quality of the evidence is low.[222,238]

There is evidence of equally low quality that a higher number of curative operations are done for recurrence when CEA is included in the follow-up rather than no CEA.[238]

When looking at the intensity of CEA testing i.e. more versus less CEA there is low quality evidence of a survival advantage with more frequent CEA measurements.[238]

There is low quality evidence that the number of recurrences detected is not affected by CEA estimation.[238]

There is also low quality evidence that the number of curative re-operations performed is greater in the group with more frequent CEA measurements.[238]

Liver imaging

There is evidence of moderate quality that liver imaging in general gives a significant survival advantage. There is also evidence of moderate quality showing no difference in the number of recurrences detected between follow-up that included liver imaging and follow-up with no liver imaging.[222]

There is evidence of low quality that ultrasonography improves survival, evidence of low quality that it increases the number of recurrences detected and evidence of low quality that it increases the number of curative re-operations. All of the above were compared with control patients in whom no ultrasonography was performed at follow-up.[238]

There is evidence of moderate quality that CT improves survival (but no statistical significance was reached), evidence of moderate quality that it increases the number of recurrences detected and evidence of low quality that it increases the number of curative re-operations. All of the above were compared with control patients in whom no CT was performed at follow-up.[238]

Colonoscopy

There is evidence of moderate quality that colonoscopic surveillance leads to a survival advantage over no colonoscopic surveillance.[238]

There is evidence of low quality showing that there is no difference in the number of recurrences detected with the addition of colonoscopy, and equally low-quality evidence showing that more curative operations for recurrence were carried out in patients who had colonoscopic surveillance during follow-up.[238]

When looking at the intensity of colonoscopy i.e. more versus less colonoscopy there is evidence of low quality that intensive colonoscopic surveillance does not offer any advantage in overall survival versus less intensive colonoscopic surveillance. [238,239]

The evidence is again of low quality that frequent colonoscopic surveillance increases the number of recurrences detected.[238,239]

Evidence of equally low quality suggests frequent colonoscopy does increase the number of curative operations attempted for recurrence.[238,239]

There is also evidence of low quality that the time to the diagnosis of a recurrence is reduced and that the survival time after recurrence is diagnosed is increased.[239]

Complications:

1 study reported adverse events from follow-up: 2 perforations and 2 GI bleeds from a total of 731 colonoscopies.[239]

Quality of life:

1 study (597 patients) reported a small but significant increase in the quality of life of patients associated with more frequent follow-up visits.[240] A different study (203 patients) reported no difference in quality of life, anxiety, depression, or patient satisfaction in patients followed up in different settings (general practice versus hospital).[241]

8.5 Discussion

These results show that overall more parameters are taken into account when assigning a quality value to a piece of evidence with GRADE compared to the sole use of internal validity checklists.

In addition all individual outcomes reported within a systematic review are graded by the GRADE method whereas internal validity checklists assess the overall quality of the systematic review methodology and its principle outcome and not each individual outcome.

In addition, these results show that in the case of this guideline topic on the follow-up of CRC this has helped highlight a number of low- and very low-quality outcome results within systematic reviews of high quality.

This is very important for guideline developing group members as it gives a clear presentation of low quality results that perhaps would have easily been mistakenly accepted as high quality only because they were presented within a review with robust methodology.

Internal validity checklists used prior to GRADE are not incorrect and indeed have a valid place in the assessment of evidence.

When assessing the quality of a study or systematic review the majority of internal validity checklists focus on the study question and the study methodology.

The NICE methodology checklists used in the initial quality assessment of this systematic review provided a thorough quality assessment of the methodology of the studies.

The checklist for the systematic reviews begins with an assessment of the review question as the first step. It would be difficult to assess how well the study has met its objectives or how relevant its conclusions are to the guideline without this.

The next item on the checklist is the description of the methodology, which is also important. One of the key distinctions between a systematic review and a general

review is the systematic methodology used. A systematic review should include a detailed description of the methods used to identify and evaluate individual studies. If this description is not present, it is not possible to make a thorough evaluation of the quality of the review.

Assessment of the literature search follows, as this needs to be sufficiently rigorous to identify all the relevant studies. A systematic review based on a limited literature search (e.g. Medline only) is likely to be heavily biased. Any indication that hand searching of key journals, or follow-up of reference lists of included studies were carried out in addition to electronic database searches is additional evidence of quality methodology.

Inclusion and exclusion criteria are also important markers of study quality and it is the next item on the checklist. A well-conducted systematic review should have used clear criteria to assess whether individual studies have been well conducted before deciding whether to include or exclude them. In addition studies covered by a systematic review should be selected using inclusion criteria that include, either implicitly or explicitly, the question of whether the selected studies can legitimately be compared. It should be clearly ascertained, for example, that the populations covered by the studies are comparable, that the methods used in the investigations are the same, that the outcome measures are comparable and the variability in effect sizes between studies is not greater than would be expected by chance alone.

As a result of the quality assessment of the two systematic reviews with the NICE checklists it was concluded that the two reviews had robust methodology. Sources of bias were identified and these were minor differences between the population of the PICO and the population of the systematic reviews and additionally the heterogeneity introduced through the sheer variability of the follow-up protocols but neither of these were considered critical in the overall quality of the systematic reviews. Therefore their results were presented in the evidence summary. The individual outcome results presented in the evidence summary did not have any further quality assessment. All the results were presented with an assumption that the quality of the systematic reviews from which they were derived transferred to all the results presented in the evidence summary.

The checklist for randomised controlled studies has a similar purpose and begins with addressing recruitment and randomisation. Random allocation of patients to treatment or no treatment is fundamental to this type of study.

The next item on the checklist is allocation concealment. This also seeks to eliminate selection bias during the process of recruitment and randomisation, and it is a marker of study quality. Centralised allocation, computerised allocation systems or the use of coded identical containers would all be regarded as adequate methods of concealment, and may be taken as indicators of a well-conducted study.

The next issue addressed is whether the treatment and control groups are similar at the start of the trial. Patients selected for inclusion in a trial should be as similar as possible, in order to eliminate any possible bias. The study should report any significant differences in the composition of the study groups in relation to gender mix, age, stage of disease (if appropriate), social background, ethnic origin or co-morbid conditions. These factors may be included in the reporting of the inclusion and exclusion criteria rather than being reported separately.

Maintaining equal management of both groups during the duration of the trial is also an important way to eliminate the introduction of bias. The only difference between the groups should be the treatment under investigation. If some patients received additional treatment, even if of a minor nature or consisting of advice and counselling rather than a physical intervention, this treatment is a potential confounding factor that may invalidate the results.

The follow-up period is equally important in the maintenance of good quality throughout the study period. The drop out rate for a study is the percentage of the individuals or clusters recruited into each treatment arm of the study who dropped out before the study was completed. The number of patients who drop out of a study should give concern if the number is very high. It is an indication of attrition bias. Some regard should be paid to why patients dropped out, as well as how many. It should be noted that the drop-out rate may be expected to be higher in studies conducted over a long period of time.

Analysis by intention to treat means that all the subjects are analysed in the groups to which they were randomly allocated. In practice, it is rarely the case that all patients allocated to the intervention group receive the intervention throughout the trial, or that all those in the comparison group do not. Patients may refuse treatment, or contra-indications arise that lead them to be switched to the other group. If the comparability of groups through randomisation is to be maintained, however, patient outcomes must be analysed according to the group to which they were originally allocated, irrespective of the treatment they actually received.

The outcomes of the study also need to be appropriate and relevant and measured in a standard, valid and reliable way.

The final issue addressed by the checklist is that of ‘blinding’. ‘Blinding’ refers to the concealment of group allocation from one or more individuals involved in a clinical research study. It seeks to reduce performance and ascertainment bias after randomisation.[244]

If bias is introduced during a trial because of differential treatment of groups or biased assessment of outcomes, no analytical techniques can correct for this limitation. Differential treatment or assessment of participants potentially resulting in bias may occur at any phase of a trial. The optimal strategy to minimize this bias is to ‘blind’ as many individuals as possible in a trial.[244]

If possible, trial researchers should ‘blind’ 5 groups of individuals involved in trials: participants, clinicians (surgeons), data collectors, outcome adjudicators and data analysts.[244]

‘Blinding’ is not an all-or-nothing phenomenon; researchers may ‘blind’ any of the involved groups. Thus, it is far preferable for researchers to explicitly state which individuals in the trial were ‘blinded’. The higher the level of ‘blinding’, the lower is the risk of bias in the study.[244]

In the RCT by Wang et al [239] there was no reference at all to methods of ‘blinding’ having been applied to any of the individuals involved in the trial and this is the only major criticism of the trial.

Randomized controlled trials of surgical interventions are frequently more difficult to ‘blind’ than RCTs of medications, which typically achieve ‘blinding’ with placebos. Whereas medical trials usually incorporate placebo medications to achieve ‘blinding’, surgical treatments often result in incisions and scars that may differ between groups.

Furthermore, if a trial aims to compare surgical therapy to non-operative management, it will often be impossible to conceal group allocation from at least some of the individuals involved in a trial (such as the patients and surgeons). Thus, surgeons must interpret the results from trials that have not been ‘blinded’ with caution.

When data collectors or outcome adjudicators cannot be ‘blinded’, researchers should ensure that the outcomes being measured are not only reliable but as objective as possible. Consideration might also be given to the use of duplicate assessment of outcomes and reporting the level of agreement achieved by the assessors.

With the use of the NICE checklist the lack of ‘blinding’ was recorded in the quality assessment of the Wang trial but the checklist does not aid quantification of this any further.

With GRADE the quality of the study and the quality of the data for each PICO outcome of the topic is used.

The GRADE-pro software generates an evidence ‘quality score’ for each outcome from responses to several specific questions about the methodological quality of the study and the data. It emphasizes outcomes.

Each of the two systematic reviews included 8 RCTs in total. They had 7 in common and 1 different.

Jeffery et al included an RCT that had quality of life as its primary outcome [241] but excluded the GILDA trial [244] as published results were on fewer patients than the number they initially set out to recruit (1240 patients rather than target of 2920) and follow-up is short (recruitment started in 1998 and the trial is still ongoing).

Tjandra et al in their meta-analysis did not include the trial looking at quality of life [241] but included the preliminary results of the GILDA trial [244]. However, the group did do a sub-group analysis excluding the GILDA results and found that the trial exclusion did not alter the overall result of their meta-analysis.

7 of the 8 RCTs in Jeffery and all of the RCTs in Tjandra had unclear reporting of their allocation concealment relevant to their randomisation process. Each outcome derived from these studies has therefore been downgraded for this reason.

Both reviews in addition to 5-year survival carried out a number of comparisons addressing different associated outcomes. Some of these are statistically significant and in others the pooled groups are either too small, and therefore the result is imprecise, or too heterogeneous and therefore the result is inconsistent.

When this information is entered into the GRADEpro software the outcome result is given a reduced score. If the outcome is afflicted by only one type of bias then the quality grade is reduced to moderate from high. If the outcome is afflicted by both imprecision and inconsistency then the quality score is low.

In the case of the survival outcome in association with the clinical setting of the follow-up the outcome results were given a quality score of ‘very low’ because there was unclear allocation concealment, there was inconsistency (heterogeneity), and serious imprecision (CI includes 1 and the total number of events was less than 300).

Using CI as a measure of precision is accepted statistical practice. The reference to the number of accrued events is however less widespread but is one of the quality items recommended through the GRADE pro software.

A low number of accrued events is considered to increase the risk of treatment effect overestimation.[245] This goes against the purpose of a trial looking alternative interventions which is to generate an estimate of effect that closely approximates the true effect and is not misleading. Harm may result from misleading findings if these are the supporting evidence for clinical guidelines. There is no clear answer as to how many events accrued is enough but a figure of between 200 and 400 has been suggested by a team of experts in this field as a reasonable rule of thumb.[245]

The number 300 is recommended by the GRADE working group as a threshold number of events below which imprecision should be considered and is included in the guidance to reviewers as part of the GRADE pro software.

Another instance of a very low quality result is that from the Tjandra review of overall survival in association with the inclusion of CEA testing in the follow-up protocol. This outcome result is methodologically interesting as both of the systematic reviews by Jeffery et al and Tjandra et al report results on this outcome but their results differ.

The two systematic reviews differ in the inclusion of the Secco RCT[243]. The inclusion of the second RCT by Secco et al in the pooled analysis by Tjandra et al is problematic. The review by Jeffery et al excluded this trial from the pooled comparison and commented that they felt survival data could not be extracted.

The review by Tjandra et al has included this trial in the analysis. The Secco publication and their results was reviewed for this analysis and the views expressed in this thesis are in agreement with the opinion expressed in the review by Jeffery et al.

From the data in the publication it is not possible to extract survival data even though the conclusion from the Kaplan Meier analysis is that there is a significant survival advantage between the groups they compare.

For the Tjandra group to have been able to pool these two studies it can only be assumed that the group wrote to the Secco group authors and received further data. However this is not mentioned in the publication. Therefore caution is recommended when considering these results

It is a common assumption that systematic reviews are always at the top of the pyramid of evidence grading. When assessing the quality of a study the majority of internal validity checklists focus on the study question and the study methodology. But in this way the emphasis is on the quality of the study itself and not on the evidence in it. So a study that is of high quality with regard to its methodology may mislead someone in thinking that all the conclusions and indeed all the evidence it is based on is also of high quality. [216]

What a GRADE analysis shows is that this might not always be the case. GRADE helps to clarify the quality of the evidence.[216]

The advantage of GRADE over other evidence grading systems is that it provides comprehensive criteria for downgrading and upgrading quality of evidence with explicit evaluation of the importance of all outcomes within the studies.[216]

This helps provide a clear separation between body of evidence and strength of recommendations, with additionally a transparent process that defines and exposes progression from the evidence to the recommendations.

The quality score provided by GRADE helps quantify the confidence in the evidence, which in turn helps guideline panels formulate transparent recommendations. If the evidence is of high quality then the recommendation might be that patients ‘should’ be treated by the method supported by the evidence. If the evidence is of lower quality then the recommendation might be that the patients ‘should probably’ be treated by this method but implying or clearly stating that this might not be appropriate in all cases.[220]

The topic on CRC follow up is complicated with many different tests and strategies featuring in a wealth of different quality therapeutic data. GRADE was shown in these results to provide an important role in providing the GDG with a clear presentation of the true quality of the different outcomes from all the studies. Despite the wealth of studies the results vary from those of moderate quality to those of very low quality and it is the latter that have serious methodological weaknesses and therefore these results need to be interpreted with caution.

Transparency should be a key goal for any guideline development group and GRADE contributes to the development of methodologically transparent guideline recommendations.

9.0 The international nature of evidence and the influence this has on the formulation of guideline recommendations

9.1 Introduction

Evidence for clinical guidelines is made up of studies from across the world. This potentially diverse geographical source of the evidence can create a concern that a clinician might be forced by guidelines to rely upon evidence that is only doubtfully relevant.[246]

Confidence might be lost in the guidance if it is thought to be formulated in the absence of evidence that is clearly applicable to the case in hand, and with evidence generated perhaps in a different category of patients in another country, at some other time, and in different circumstances.[246]

This would be far from ideal and indeed it might be considered to be biased-medicine rather than evidence-based.[246]

It is also possible that different guideline developers might use the same body of evidence to produce different guidelines for a different population or setting.

It is vital that recommendations with any guideline are clear, and are based on the best available evidence.

Interpreting the evidence to make recommendations is at the heart of the work of the Guideline Development Group (GDG). It is not a straightforward task and it is challenging.[54]

The GDG must decide what the evidence means in the context of the review questions and the economic questions posed, and decide what recommendations can usefully be made to healthcare professionals.[54]

In a NICE guideline an important aim is to show clearly how the GDG moved from the evidence to the recommendation. This is best done in a section called ‘linking evidence to recommendations’ (LETR).[54]

This section may also be a useful way to integrate the findings from several evidence reviews that are related to the same recommendation(s).[54]

In this section the GDG’s view of the strength of a recommendation should be clearly reported. Points to address include the prioritisation of outcomes, the trade-off between clinical benefits and harms, the trade-off between net health benefits and resource use and the quality of the evidence.[54]

With reference to the latter in particular, there should be discussion of how the presence of potential biases and uncertainty in the clinical and economic evidence has influenced the recommendation, and why.[54]

This may include consideration of whether the uncertainty is sufficient to justify delay in making a recommendation to await further research, taking into account the potential harm of failing to make a clear recommendation.[54]

9.2 Aim

To research the international nature of data supporting the NICE guideline on CRC and the influence this may have on the resulting recommendations.

9.3 Methods

Information on how the evidence that supported the guideline was linked to the final recommendations formulated by the guideline development group was collected through:

i) participation in all the guideline development group (GDG) meetings throughout the two year development of the CRC guideline and experiencing in close proximity the process of developing recommendations from the body of evidence presented for each topic.

ii) review of all the final formal LETR sections of the CRC guideline document for each of the topics the guideline addressed for evidence of the recommendations having been influenced by the national setting of the evidence that supported that recommendation.

9.4 Results

Of the fifteen topics and questions which the NICE colorectal cancer guideline addressed there was only one that was found to be a demonstration of a guideline recommendation potentially influenced by the national setting and origin of the supporting evidence. This guideline topic was on the staging of CRC and is presented below.

The summary of the systematic review evidence is presented as part of this thesis as it is necessary to the understanding of the topic and the detailed evidence that was presented to the guideline development group before they made their assessment and formulated the recommendation. The author of this thesis did not perform this systematic review. It is the work of colleague reviewer Dr Susan O'Connell, permanent staff of the NCC-C technical team and member of the NICE CRC guideline development group.

The objective of this thesis chapter is not the exercise of performing a relevant systematic review but instead the critical appraisal of the process that links the evidence of a systematic review to the formulated recommendations and the analysis of the consequences and implications.

8.4.1 Guideline topic identified demonstrating influence of the recommendation by the national setting of the supporting evidence

Guideline question [247]

"For patients diagnosed with primary colorectal cancer, what is the most effective technique(s) in order to accurately stage the disease (excluding pathology)?"

Evidence summary for rectal cancer [247]

"From two systematic reviews (Kwok et al.[248], Bipat et al.[249]) it appears that endorectal sonography/endorectal ultrasound had the highest sensitivity, specificity and accuracy of the modalities investigated (CT, endorectal sonography/endorectal ultrasound and MRI)." [247]

"Kwok et al. [248] reported a pooled sensitivity, specificity and accuracy for endorectal sonography of 93%, 78% and 87% respectively for wall penetration and 71%, 76% and 74% respectively for nodal involvement." [247]

Bipat et al. [249] reported summary estimates of sensitivity and specificity for endorectal ultrasound of 94% and 86% respectively for muscularis propria invasion, 90% and 75% respectively for peri-rectal tissue invasion and 67% and 78% respectively for lymph node involvement compared with sensitivity and specificity for MRI of 90% and 69% respectively for muscularis propria invasion, 82% and 76% respectively for peri-rectal tissue invasion and 66% and 76% respectively for lymph node involvement. [247]

"For muscularis propria invasion, endorectal sonography specificity was significantly higher than that of MRI (p=0.02); for peri-rectal tissue invasion, endorectal ultrasound sensitivity was significantly higher than that of CT (p<0.001) and MRI (p=0.003)." [247]

"Specific UK evidence was provided from the Mercury Study group [250,251] investigating MRI in the staging of rectal cancer. The accuracy of MRI for predicting the status of circumferential resection margin (presence/absence of tumour) by initial imaging or imaging after pre-operative treatment was 88% [95% CI: 85-91%], sensitivity was 59% [95% CI: 46-72%] and specificity was 92% [95% CI: 90-95%]." [247]

"For patients undergoing primary surgery with no pre-operative treatment (n=311), accuracy of prediction of a clear margin was 91% [95% CI: 88-94%], sensitivity of 42% and specificity of 98%." [247]

"For patients undergoing pre-operative chemoradiotherapy or long-course radiotherapy the accuracy of prediction of clear margins on MRI was 77% [95% CI: 69-86%], sensitivity was 94% and specificity was 73%." [247]

"Two studies investigated the use of FDG PET (Kantorova et al [252], Llamas-Elvira et al [253]." [247]

"For lymph node involvement the reported sensitivity ranged from 21-29%, specificity ranged from 88-95% and accuracy ranged from 56-75% and for liver involvement sensitivity was 78%, specificity was 96% and accuracy was 91%." [247]

"Inter-observer agreement was not addressed in all studies, though the studies which did evaluate inter-observer agreement (Fillipone et al.[254], Tatli et al.[255], Kim et al. [256]) reported good to excellent agreement for interventions being investigated." [247]

<u>Recommendations for staging of rectal cancer</u> [247]

"Offer contrast enhanced CT of the chest, abdomen and pelvis, to estimate the stage of disease, to all patients diagnosed with colorectal cancer unless it

is contraindicated. No further routine imaging is needed for patients with colon cancer." [247]

"Offer magnetic resonance imaging (MRI) to assess the risk of local recurrence, determined by anticipated resection margin, tumour and lymph node staging, to all patients with rectal cancer unless it is contraindicated." [247]

"Offer endorectal ultrasound to patients with rectal cancer if MRI shows disease amenable to local excision or if MRI is contraindicated." [247]

"Do not use the findings of a digital rectal examination as part of the staging assessment." [247]

Linking evidence to recommendations [247]

"The GDG placed a high value on accurate staging at presentation because this information informs the optimum treatment strategy for patients with colorectal cancer. The evidence consisted of two good quality systematic reviews and several low-quality case series studies." [247]

"In patients with rectal cancer, the GDG were aware that the available evidence had shown EUS to have higher sensitivity, specificity and accuracy compared to MRI or CT for identifying those patients whose tumours are suitable for local resection. The GDG noted that EUS is not appropriate in bulky, obstructing tumours and does not visualise the total extent of nodal disease in the pelvis." [247]

"It was also noted that the evidence may reflect non-UK practice because EUS is not widely used in the UK." [247]

"There was also significant inter-observer variation in the performance of EUS." [247]

"The GDG therefore recommended MRI be used for the initial assessment of patients with rectal cancer and that EUS be considered if the MRI suggested disease which was amenable to local resection." [247]

8.4.2 The national setting of the studies included in the evidence summary

The national setting of the studies included in the evidence is presented in Table 9.1.

Table 9.1: The national setting of studies included in the evidence summary for the guideline topic on the staging of rectal cancer

Study (first author and date)	National setting
Kwok [248]	New Zealand
Bipat [249]	The Netherlands
Mercury Study Group [250,251]	UK
Kantorova [252]	Czech
Llamas-Elvira [253]	Spain
Fillipone [254]	Italy
Tatli [255]	USA
Kim [256]	Korea

The two systematic reviews by Kwok [248] and Bipat [249] pooled 83 and 90 studies respectively. The Kwok group was based in New Zealand and the Bipat group in The Netherlands.

For the two systematic reviews Table 9.2 presents the national setting of the individual studies pooled in the reviews.

Table 9.2: The national setting of the studies pooled by the two systematic reviews and included in the evidence summary of the guideline topic on the staging of CRC cancer

.

National setting of study	Number of studies from Bipat review [249]	Number of studies from Kwok review [248]
UK	7	4
Austria	1	1
Germany	11	11
Italy	17	10
Sweden	5	6
Finland	1	1
France	2	1
The Netherlands	2	1
Switzerland	1	1
Denmark	3	3
Ireland	1	0
Belgium	0	1
Croatia	0	1
Russia	1	1
USA	20	19
Japan	11	12
Australia	1	2
Korea	4	2
Canada	0	1
Total	88 *unable to obtain data for 2 studies	78 * unable to obtain data for 5 studies

9.5 Discussion

These results show that the GDG made a recommendation that was in part influenced by the national setting of the evidence which emerged from the systematic search of the literature.

The results of the literature search identified two high quality systematic reviews which concluded that the modality that was shown to have the highest diagnostic accuracy in staging rectal cancer was endoanal ultrasound.

This accuracy was higher than that of the MRI scan. This was clearly acknowledged by the GDG in the relevant LETR paragraph.

Despite this the GDG made a decision to recommend that patients with rectal cancer in the UK are offered an MRI to stage rectal tumours and that an endoanal ultrasound is to be offered only to those patients in whom MRI has shown the tumour to be amenable to local excision or where MRI is contraindicated.

One of the reasons given in the LETR paragraph for this decision was the fact that endoanal ultrasound is ‘non-UK practice’.

It appears that the UK Mercury study [250, 251], a high quality study researching the diagnostic accuracy of MRI in relation to the staging of rectal cancer, was upgraded in the evidence pool by the GDG despite the fact that in absolute terms the MRI results it presented did not show MRI to be a more accurate diagnostic modality compared with EUS.

Though this decision is not entirely based on the national setting of a particular study (the GDG also considered the low inter-observer variability as a reason for their decision) this case is an example where the national setting of evidence does play a significant role in the formulation of their decisions.

This may be the case because the difference in diagnostic accuracy between the modalities is actually rather small. Had it been a more remarkable difference in diagnostic accuracy between EUS and MRI then perhaps it would have been more difficult to overlook a result of diagnostic accuracy and give preference to the option that had local origins.

Another reason which may have lead the GDG to show preference to local results is the opinion that the results of the Mercury study [250, 251] are more applicable and transferable to the population that the guideline intends to serve.

The results are more applicable because of the characteristics of the study population, which sampled patients from the local UK population. The results are also more applicable because they have been produced by modalities with technological characteristics similar to those modalities used regularly in UK clinical practice.

In addition, there has been a long and concentrated effort in the UK to standardise the performance and reporting of MRI scans. Dr Gina Brown, Consultant Radiologist at the Marsden Hospital has produced a lot of detailed work in standardising scanning techniques and has produced a detailed proforma so that all radiologists nationally can report their findings including all the relevant details. This is to give every possible chance that the diagnostic accuracy reported in trials like the Mercury trial are replicated nationally. [259] The same is not currently available for EUS. The combination of the reduced availability of specialists to perform the test confidently coupled with the absence of standardised ways to perform and report the scans could result in an inability to achieve similar results to the trial.

In this way the GDG may also have downgraded evidence from foreign national settings due to an opinion that those studies were carried out within health systems very different from that in the UK.

The US is an example where the health system is in the majority private, EUS is performed by surgeons rather than radiologists, and payment for the service is arranged differently (pay for service). A large proportion of studies included in the systematic reviews by Bipat [249] and Kwok [248] were performed in the US.

These are also differences found to a lesser degree within countries of the European Union where the presence of national health systems coexists with a variable degree of private health provision along with variable access policies to either of these. This can create very different working environments and research environments which influence the results of foreign studies and the evidence that can be drawn from them.

In addition there are many cultural differences with regard to health and the investigations and treatment of ill-health across the world. The example of the EUS modality lends itself to demonstrate this due to its invasive nature and per rectum administration.

Some populations are more accepting of certain modes of treatment that others may find too invasive to tolerate and would rather reserve such treatment as the last resort in the investigation or treatment pathway. It is perhaps an expression of cultural differences through health experiences.

This may perhaps be a subtle reason why EUS was less favoured among a UK guideline development group if the investigation was considered a less favourable option by the population the GDG members are used to treating.

The members of the guideline development group know that MRI is more available than EUS in the UK. They are also aware that the available expertise for the performance of the two tests is very different in the UK. MRI interpretation is readily available by specialist radiologists whereas the performance and interpretation of EUS is only available from a small number of specialist radiologists in tertiary care hospitals.

There is a chance that the members of the GDG have made their decision in order not to disrupt the ‘status quo’ of current service provision.

There is also the possibility that certain members of the GDG have a conflict of interest that has not been taken into account. There are members of the GDG who have specialist interest in imaging and this representation may be stronger than that regarding interest in EUS.

Any conflict of interest of members of GDG is reported and recorded as part of the NICE guideline development methodology and there is a policy of transparency with regard to this matter. However this is usually regarding active participation in ongoing trials or links with industry and this more subtle potential conflict of interest may be overlooked.

There is also a simple sense of familiarity with results of local research which, for the members of the GDG, may be difficult to escape when making their decision; and so when a study conducted locally offers slightly less accurate diagnostic efficacy from a modality but which gives results that are not very different from those produced in other national settings the GDG members have an automatic preference for what is familiar.

Guidelines are produced to improve medical practice and patient care. In a study by the World Health Organisation on identifying barriers to the implementation of guidelines it was identified that physicians, in addition to quality evidence, also need local evidence in order effectively to change their practice. This is because they need to validate the transferability of the findings in their own practice.[257]

There are methodological implications that follow from the decision of the GDG however since any of the reasons presented above may introduce bias to the process.

Upgrading or downgrading evidence based on the national setting of the evidence is not incorrect however and at times it is necessary. It requires skill

and experience from members of a GDG panel in order to make accurately the right recommendation for the relevant population. Because the decision can expose the process to the risk of bias it is important that the process itself is by design as transparent as possible.

It must be clear at all times what the evidence has shown and how this has been interpreted by the particular GDG.

Theoretically guideline recommendations should be based on the best available evidence. They should be free from financial constraints and be the result of a pure methodological process.

The results have shown that recommendations are not always pure and without influence. They are pragmatic decisions about the UK population within the framework of the NHS. They are however based on the best evidence available. Especially when the differences in the data between studies are not very substantial and management recommendations are potentially controversial, then decisions can become more subjective than objective and subject to the influence of the national setting of the studies under consideration.

Provided that there is a process such as the ‘linking evidence to recommendation’ section of the NICE guidelines so that all readers and users of the guideline know how the evidence was interpreted, then influence on the recommendations by the national setting of the evidence under review is not a weakness but might even be seen as a necessary approach towards the proper assessment of the vast quantities of evidence reported in the literature.

10.0 Conclusions and future work

The results of this thesis show that guideline development at NICE follows an evidence-based process that complies with internationally accepted principles of evidence-based medicine and guideline development.

This process is continually reviewed and updated in order to ensure quality and transparency. It is well regarded both nationally and internationally for these reasons.

No guideline development process is free of bias and some steps of the NICE guideline development process are more robust than others.

One such example is that the technical team assigned to the development of each NICE guideline includes only one reviewer who performs the systematic reviews.

It is strongly advised in systematic review methodology literature [260] that reviewing is carried out by at least two reviewers. They work independently to screen abstracts, extract data and assess risk of bias, thereby reducing the chance of reviewer bias and increasing reliability.

The most likely reasons for NICE to divert from this principle of guideline methodology are time and cost constraints.

The NICE colorectal cancer guideline took two and a half years to develop which is average for NICE clinical guidelines.

This is a lengthy process considering that information is made available at such speed and with such ease of access. Guidance is therefore often desired at a fast pace.

Guidelines are also gaining popularity and developers are facing a huge expansion in their output.

Therefore, the cost of the development process for each guideline must be balanced against the need and resulting costs of producing a larger number of guidelines.

It is understandable that a guideline development body desires to reduce both the development time and the costs.

However, this is such an important factor in reducing reviewing bias that it would be advisable that the methodology specialists at NICE address this in the future.

It is especially important as NICE want to be regarded as leaders in the field of guideline development.

It is also important as many other countries use NICE guidance and NICE guideline methodology as a template for their own guideline development.

This is either directly through the services of NICE International or by following NICE methodology in the set up of their own guideline developing bodies.

Most especially it is important because reducing reviewing bias increases the reliability of the results and strengthens the resulting recommendations.

This thesis has also shown that there is variability in the availability and quality of the evidence that underpins the NICE guideline and this has the potential of introducing further bias.

Transparency is imperative at all levels of the development process so that there can be no confusion about the quality of the results and the strength of the recommendations.

The NICE guideline development process is transparent. The methodology is clear and easily available. Conflicts of interest are declared by the members of the guideline development groups.

In addition, the formulation of the recommendations by the panel of experts is explicitly presented in the ‘linking evidence to recommendations’ paragraph at the end of each recommendation.

However it is important to understand and highlight that a guideline is not a regulation, nor a statuatory obligation. It is the opinion of the committee that formulates the recommendations based on the evidence provided to them.

This evidence may be biased, interpreted under certain influences, or simply lacking as demonstrated by the work in this thesis. When this is the case then the recommendations must be presented and accepted by all as being weak.

Even in the face of strong evidence, and a strong recommendation it is also important for clinicians to understand that a guideline is still guidance and not an over-arching inflexible obligatory rule that they need to adhere to in fear of being accused of professional misconduct.

This is particularly important as NICE has currently been given the remit by the department of Health and NHS England to develop quality standards.

NICE quality standards are concise sets of prioritised statements designed to drive measurable quality improvements within a particular area of health or care.[261]

They are derived from the best available evidence such as NICE guidance and other evidence sources accredited by NICE. They are developed independently by NICE.[261]

NICE works with independent Quality Standards Advisory Committees (QSAC) to develop the quality standards. Standing members are drawn from

the NHS, health, public health and social care professionals, patients/service users and carers and academia.

The members of the QSAC do not represent their organisations but are selected for their expertise, experience of working with multidisciplinary and lay colleagues and understanding of evidence based care. A number of topic experts are invited to join the standing members for each quality standard topic.[261]

The aspiration is that the quality standards will enable health, public health and social care practitioners to make decisions about care based on the latest evidence and best practice.[261]

There is also the aspiration that the standards will help people receiving health and social care services, their families and carers and the public in general to find information about the quality of services and care they should expect from their health and social care provider.[261]

It is also hoped that service providers will refer to the standards in order to quickly and easily examine the performance of their organisation and assess improvement in standards of care they provide.[261]

Commissioners might also be able to refer to the standards so that they can assess whether the services they are purchasing are high quality and cost effective and focused on driving up quality.[261]

However, the term ‘standard’ gives the impression of something more binding than guidance. This creates the anxiety that what started as guidance is converted into something that practitioners feel less able to deviate from.

Considering the finding of this thesis, that the evidence underpinning NICE guidance is varied in its availablity and quality, coupled with the knowledge that NICE guidelines are the primary evidence undepinning the development of the standards, gives reason for concern.

There are two main components to a quality standard: the quality statement and the quality measure.[261]

The quality statement generally describes high-priority areas for improvement. They are aspirational but achievable. Each statement specifies one concept or requirement for high quality care or service provision, for example a single intervention, action, or event.[261]

The quality measure that accompanies every statement is intended to be able to assess the quality of care or service provision specified in the statement.[261]

NICE quality standards are not mandatory nor are they targets. The audience of these quality standards needs to be aware of this fact and therefore approach these standards accordingly.[261]

Future work in understanding the development of the quality standards and particularly how NICE guidance is translated into a NICE quality standard is required.

A specific area of research should include how each guideline recommendation is converted to a standard.

It would be of interest to explore whether all recommendations are converted into a standard and how the strength of the recommendation is evident in the resulting standard.

In addition it would be interesting to research how transparent the methodology relating to the development of quality standards is.

Particularly looking at the reporting of bias in the evidence and how this remains transaparent in the developing standard.

The results of this thesis have also shown that there is poor reporting of methodological data by many of the UK surgical societies and Royal Colleges which produce guidelines.

Though this is not an indication that these guidelines are of poor quality, it does prohibit their appropriate appraisal by their users. This is another area where future work can be done so improvements can be made.

Appropriate attention must be given by professional societies and Royal colleges to the online publication of methodology data.

Ideally this should be overseen by guideline development specialists if they havent already been involved in the development process.

However, the results of this thesis demonstrate that there are many specialist considerations that are involved in guideline development.

Even if the results of this thesis purely represent a failure of online presentation of methodology data it has been demonstrated that this is a highly specialist discipline that demands both time and resources so that the resulting guidelines can be of high quality.

There are a variety of tools such as GRADE, and AGREE that are the products of lengthy research in this discipline. These assist in the development and appraisal of the guidelines.

The quality assessment of therapeutic study data with GRADE in this thesis highlighted the importance of assessing all outcomes reported within a study.

It is a common assumption that systematic reviews are at the top of the pyramid of evidence quality. These results have shown that it is possible to have some very low quality outcome results within a systematic review of high quality.

If this is not identified in a systematic fashion then it may be mistakenly overlooked leading to potentially inaccurate recommendations.

A deep sense of responsibility towards the principles of evidence based medicine, enthusiasm and resources are required in equal measures in order to be able to keep up with such developments in this field.

If the professional bodies want to be guideline producing bodies of quality then appropriate resources will need to be directed to this endeavour, perhaps more than what has been devoted to in the past.

Considering the time and resources that go into guideline production perhaps it might be a future consideration of professional societies and Royal Colleges not to engage in the development of guidelines in the same way that thay have been doing so in the past.

Rather than developing guidelines directly a better investment of resources might be the stringent appraisal of the guidelines produced by NICE.

In addition, it may be better value to invest time and resources in a more active engagement with the stakeholder process of the NICE guidelines.

Stakeholders have the opportunity through the NICE development process to engage and influence the guideline production from its infancy.

This perhaps is currently not utilised until later stages in the guideline development when it is much more difficult to influence the course of a particular guidance.

This may mean that there is less chance of contradictory guidelines being produced by guideline producing bodies.

It may also be more effective for these organisations to invest in the education of their members with regards to guideline appraisal.

Guidelines are increasing in popularity and a larger number of them are being produced both nationally and internationally.

It is important for medical practictioners to know what to look for in guidelines and their methodology in order to assess the strength of the recommendations and whether they warrant changing clinical practice.

They are more likely to look to their professional bodies and the Royal Colleges for such education which at present does not include education in guideline appraisal.

In addition, with the emergence of quality standards it will become even more important to be able to assess evidence and guidelines so as to be able to further decide whether the quality standards are indeed quality markers and not transcripts of weak guideline recommendations.

Another finding of this thesis was that a lot more investment is required in the field of data storage and analysis.

In attempting to improve treatments and therefore survival from cancer it is important to have all possible information available and easily accessible for analysis.

The regional cancer registries collect information and have quality control measures in place. There are a variety of other regional and national databases that collect and store patient data. Some have voluntary data collection which poses data quality challenges.

The improvements made in terms of the setting up of NCIN and the national registration database in England is a step in the right direction but so much more needs to be done in this field.

Though there have been important improvements in the collection and handling of patient data and particularly clinical practice data over the last five

years it is quite clear from the results of this thesis that accessing data is complicated and remains too slow.

This is not acceptable and significant further work and resources need to be directed towards this area.

The results of this thesis have identified poor methodological quality in diagnostic accuracy studies despite a wealth in the number of studies available.

This is just one example where repeatedly there is an expense of patients time and effort in addition to the cost of running trials without a satisfactory result.

The guideline was unable to make a recommendation on the topic of imaging for heaptic colorectal metastases despite a number of studies being available.

Linkages among multiple databases are increasingly used to merge clinical data with administrative data to provide the power and depth of data needed to address clinical research questions extended through time.

Linking these large databases to clinical research data can also be used to study outcomes such as health care utilisation or survival over long periods of time without having to track individual study participants themselves.

Doing so requires that critical issues related to privacy and protection of data and human research subjects be addressed.

Further potential impedements to such linkages exist including availability and accuracy of variables that could be used to correctly link an individual’s record in one database to other databases, and the technology available to perform complex data linkages.[262]

However, when this is carried out successfully, then the result is significant. The data does exists and is stored in the UK. What is needed is to be able to have the adequate technology and staff to adequately ‘clean’ the data from parameters that could bias the linkage and actually perform the linkage of the databases.

The result output is potentially large, it does not involve patients being exposed to additional risk, and the cost will compare favourably to running patient studies and trials for many years.

Guideline recommendations often conclude that further research is required in order to answer a particular clinical question. Further research need not mean traditional clinical trials only.

It is important that linkage studies are used to provide answers to guideline questions. In the same way it is important that guidelines and their need for evidence-based answers is used as a drive for improvement in this chrinically under-resourced and under-invested area of health research.

Over the years guidelines have become a common feature of medical practice and many organisations are producing them.

Their purpose is to improve the quality of healthcare provision and to create and maintain standards of care.

It is the over-arching conclusion of this thesis that guideline production through systematic searching of the evidence and with a transparent presentation of the results and strength of the recommendations is an outstanding addition to the high quality practice of modern medicine.

It must however be used with appropriate training in guideline appraisal and confidence that it is a guidance that can be subject to a variety of bias and cannot and should not dictate clinical judgment and practice in totality.

References

1. Boyle P, Langman J, ABC of colorectal cancer. BMJ. 2000;321:805-8
2. Cappell MS. The Pathophysiology, clinical presentation, and diagnosis of colon cancer and adenomatous polyps. Med Clin N Am. 2005;89:1-42
3. Markowitz SD, Bertagniolli MM. Molecular basis of colorectal cancer. N Engl J Med. 2009;361:2449-60
4. Quinn, M.Registrations of cancer diagnosed in 1994-1997, England & Wales. Health Statistics Quarterly. 2000;07:71-82.
5. Meyerhardt JA, Niedzwiecki D, Hollis D. Association of dietary patterns with cancer recurrence and survival in patients with stage III colon cancer. JAMA. 2007;298:754-764
6. Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. American Journal of Clinical Nutrition. 2007;86(3):556-565.
7. Meyerhardt JA, Giovannucci EL, Holmes MD. Physical activity and survival after colorectal cancer diagnosis. J. Clinical Oncol. 2006;24(22):3527-34
8. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and Colorectal Cancer - a meta-analysis. JAMA. 2008;300(23):2765-2778.
9. St John DJ, McDermott FT, Hopper JL. Cancer risk in relatives of patients with colorectal cancer. Ann Intern Med. 1993;118(10):785-90.
10. Veale AMO. Intestinal polyposis. In: Eugenis Laboratory Memoirs Series 40. New York:Cambridge University Press; 1965.p.225-239
11. Cockyne EA. Hereditary in relation to cancer. Cancer Rev. 1967;2:337–47.
12. Herrera L, Kataki S, Gibas L, Pietrzak E, Sandberg AA. Gardner syndrome in a man with an interstitial deletion of 5q. Am J Med Genet 1986;25:473–6.

13. Bodmer WF, Bailey CJ, Bodmer J, Bussey HJ, Ellis A, Gorman P, et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature 1987;328:614–6.
14. Leppert M, Burt R, Hughes JP, Samowitz W, Nakamura Y, Woodward S, et al. Genetic analysis of an inherited predisposition to colon cancer in a family with a variable number of adenomatous polyps. N Engl J Med 1990;322:904–8.
15. Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy DB, et al. Identification of FAP locus genes from chromosome 5q21. Science 1991;253:661–5.
16. Olschwang S, Laurent-Puig P, Groden J, White R, Thomas G. Germ-line mutations in the first 14 exons of the adenomatous polyposis coli (APC) gene. Am J Hum Genet 1993;52:273–9.
17. Jen J, Powell SM, Papadopoulos N, Smith KJ, Hamilton SR, Vogelstein B, et al. Molecular determinants of dysplasia in colorectal lesions. Cancer Res 1994;54:5523–6.
18. Miyaki M, Konishi M, Kikuchi-Yanoshita R, Enomoto M, Igari T, Tanaka K, et al. Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. Cancer Res 1994;54:3011–20.
19. Suraweera N, Duval A, Reperant M, Vaury C, Furlan D, Leroy K, et al. Evaluation of tumor microsatellite instability using five quasimonomorphic repeats and pentaplex PCR. Gastroenterology 2002;123:1804–11.
20. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N Engl J Med 1988;319:525–32.
21. Robbins DH, Itzkowitz SH. The molecular and genetic basis of colon cancer. Med Clin North Am 2002;86:1467–95.
22. Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM, et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science 1989;244:217–21.
23. Slebos RJ, Rodenhuis S. The ras gene family in human non-small-cell lung cancer. J Natl Cancer Inst Monogr 1992;13:23–9.

24. Yashiro M, Carethers JM, Laghi L, Saito K, Slezak P, Jaramillo E, et al. Genetic pathways in the evolution of morphologically distinct colorectal neoplasms. Cancer Res 2001;61:2676–83.
25. Ivanovich JL, Read TE, Ciske DJ, Kodner IJ, Whelan AJ. A practical approach to familial and hereditary colon cancer. Am J Med 1999;107:68–77.
26. Chung DC, Rustgi AK. DNA mismatch repair and cancer. Gastroenterology 1995;109:1685–99.
27. Brenner DA. Gastrointestinal basic science 2002–2003: the year in review. Clin Gastroenterol Hepatol 2004;2:9–13.
28. Burri N, Shaw P, Bouzourene H, Sordat I, Sordat B, Gillet M, et al. Methylation silencing and mutations of the p14ARF and p16NK4A genes in colon cancer. Cancer Lab Invest 2001;81:217–29.
29. Shannon BA, Lacopetta BJ. Methylation of thehMLH1,p16, and MDR1genes in colorectal carcinoma: associations with clinicopathological features. Cancer Lett 2001;167:91–7.
30. Church JM. Clinical significance of small colorectal polyps. Dis Colon Rectum 2004;47:481–5.
31. Nivatvongs S. Surgical management of malignant colorectal polyps. Surg Clin North Am 2002;82:959–66.
32. Deans GT, Patterson CC, Parks TG, Spence RA, Heatley M, Moorehead RJ, et al. Colorectal carcinoma: importance of clinical and pathological factors in survival. Ann R Coll Surg Engl 1994;76:59–64.
33. Kanazawa T, Watanabe T, Kazama S, Tada T, Koketsu S, Nagawa H. Poorly differentiated adenocarcinoma and mucinous carcinoma of the colon and rectum show higher rates of loss of heterozygosity and loss of E-cadherin expression due to methylation of promoter region. Int J Cancer 2002;102:225–9.
34. Gabriel WB, Dukes C, Bussey HJR. Lymphatic spread in cancer of the rectum. Br J Surg 1935;23:394-413
35. https://cancerstaging.org/references-tools/quickreferences/Documents/ColonSmall.pdf Accessed Mar 2014

36. Fisher ER, Sass R, Palekar A, Fisher B, Wolmark N. Dukes' classification revisited:findings from the national surgical adjuvant breast and bowel projects (protocol R-01).Cancer 1989;64:2354–60.
37. Glasziou P, DelMar C, Salisbury J. What is evidence based practice. In Evidence-based practice workbook. London: Blackwell Publishing for BMJ Books; 2007.p.3-5
38. Cochrane AL. Effectiveness and Efficiency:Random Reflections on Health Services. London: Nuffield Provincial Hospital Trust;1972.p83 (reprinted in 1989 in association with the British Medical Journal).
39. http://www.cochrane.org/aboutus [Accessed 9 Feb 2014]
40. Sacket DL, Strauss SE, Richardson WS, Rosengerg W, Hayne RB. Evidence-based medicine: how to practice and teach EBM. 2nd Edition. Edinburgh: Churchill Livingstone; 2000. p.1-3
41. Institute of Medicine Committee to Advise the Public Health Service on Clinical Practice Guidelines. Definition of key terms. In: Field MJ, Lohr KN, editors. Clinical Practice Guidelines: Directions for a new programme. Washington DC: National Academy Press;1990.p.33
42. Edelstein L. The Hippocratic oath: text, translation and interpretation. Baltimore: John Hopkins University Press; 1996. p.56
43. http://www.gmc-uk.org/licencing/index.asp Accessed 3 Jul 2011
44. http://www.gmc-uk.org/good_medical_practice/index.asp Accessed 3 Jul 2011
45. Jones, M. Medical Negligence. 3rd Edition. London: Sweet & Maxwell; 2003. p 34
46. http://www.nice.org.uk/aboutnice. Accessed 5 Jul 2011
47. http://www.sign.ac.uk/about/index.html Accessed 5 Jul 2011
48. http://www.sign.ac.uk/methodology/index.html Accessed 5 Jul 2011
49. http:/ http://www.gain-ni.org Accessed 5 Jul 2011
50. http://guidelines.gov Accessed 5 Jul 2011
51. http://g-i-n.net Accessed 5 Jul 2011
52. http://www.aezq.de Accessed 5 Jul 2011
53. http://www.en.wikipedia.org/wiki/Health-economics. Accessed 28 Jul 2013

54. http://www.nice.org.uk/niceMedia/pdf/GuidelinesManualChapter8.pdf Accessed 28 Jul 2013
55. http://en.wikipedia.org/wiki/Incremental_cost-effectiveness_ratio Accessed 28 Jul 2013
56. Folland S, Goodman AC, Stano M. Economic Effiencey and Cost-Benefit Analysis. In: The Economics of Health and Health Care. 6th Edition. Boston, MA: Prentice Hall; 2010. p. 78
57. Orszag PR, Ellis P. Addressing rising health care costs—A view from the Congressional Budget Office. N Engl J Med, 2007; 357:1885–1887.
58. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, Cook J, Glick H, Liljas B, Petitti D, Reed S. Good research practices for cost-effectiveness analysis alongside clinical trials: The ISPOR RCT-CEA task force report. Value in Health, 2005; 8(5):521-533.
59. http://www.nice.org.uk/media/231/CB/NICECitizensCouncilDepartingThresholdFinal.pdf Accessed 5 Feb 2014
60. Greenhalgh T. How to read a paper. London: BMJbooks; 2001. p. xv
61. Greenhalgh T. How to read a paper. London: BMJbooks; 2001. p139-140
62. Godlee F. EBM: flawed but still the best we've got. BMJ 2014;348:440
63. Lenzer J. Why we cant trust clinical guidelines. BMJ 2013;346:3830
64. http://en.wikipedia.org/wiki/Methodology Accessed 28 Jul 2013
65. http://www.agreetrust.org/about-agree/introduction0/ Accessed 10 Feb 2013
66. http://www.agreetrust.org/about-agree/introduction0/purpose-of-agree-ii/ Accessed 10 Feb 2013
67. http://www.agreetrust.org/about-agree/introduction0/which-guidelines-can-be-appraised-with-the-agree-ii0/ Accessed 10 Feb 2013
68. Kerwin AJ, Haut ER, Burns JB, Como JJ, Haider A, Stassen N, Dahm P, and Eastern Association for the Surgery of Trauma Practice Management Guidelines Ad Hoc Committee. The Eastern Association of the Surgery of Trauma approach to practice

management guideline development using Grading of Recommendations, Assessment, Development, and Evaluation(GRADE) methodology. J Trauma Acute Care Surg. 2012;73: S283-S287.

69. http://www.nice.org.uk/media/615/64/The_guidelines_manual_2009.pdf Accessed 15 Apr 2009
70. http://publications.nice.org.uk/the-guidelines-manual-pmg6 Accessed 9 Feb 2014
71. Hooper J, Longworth P. Health needs assessment workbook. London: Health Development Agency;2002. p. 5 (www.hda.nhs.uk/publications - accessed 24 March 2010).
72. http://en.wikipedia.org/wiki/Epidemiology_of_cancer. Accessed 15 Aug 2013.
73. Dos Santos Silva I. The role of cancer registries. In: Cancer Epidemiology:Principles and Methods.Lyon:international agency for research in cancer;1999.p385-403. www.iarc.fr/en/publications/pdfs-online/epi/cancerepi/cancerepi-17.pdf Accessed 11 Feb 2014.
74. Skeet RG. Quality and quality control. In: Cancer registrations-principles and methods.Jensen OM, Parkin DM, MacLennan R, Muir CS (editors). Lyon: International agency for research in cancer and the International association of cancer registries;1991.p101-107. www.iarc.fr/en/publications/pfds-online/epi/sp95/sp95-chap9/pdf Accessed 12 Feb 2014
75. http://www.ukacr.org/content/about-ukacr Accessed 14 Feb 2014
76. http://www.ukacr.org/content/data-quality Accessed 14 Feb 2014
77. http://www.ukacr.org/registration-organisation Accessed 16 Feb 2014
78. http://www.ons.gov.uk/ons/about-ons/who-we-are/overview/index.html Accessed Apr 2009
79. http://www.ons.gov.uk/ons/about-ons/what-we-do/what-we-do-overview/index.html Accessed Apr 2009
80. http://www.ons.gov.uk/ons/publications/index.html Accessed Feb 2014
81. http://www.ncin.org.uk/about_ncin/ Accessed Jul 2011

82. http://www.ncin.org.uk/collecting_and_using_data/ Accessd Feb 2014
83. http://www.ncin.org.uk/collecting_and_using_data/national_cancer_data_repository/ Accessed Feb 2014
84. http://www.ncin.org.uk/cancer_information_tools/eatlas/ Accessed Feb 2014
85. http://www.ncin.org.uk/Prevalence/1_5_10_Year/atlas.html Accessed Feb 2014
86. http://www.cancerresearchuk.org/about-us Accessed Jul 2011
87. Office for National Statistics, Cancer incidence, mortality and survival, 2005. Accessed 2008
88. Welsh Cancer Intelligence and Surveillance Unit, , Cancer Incidence in Wales 1992-2001 December 2002. Accessed 2008
89. ISD Online Information and Statistics Division, , Lifetime risk of developing cancer, Scotland: 2001-2005 Accessed 2008
90. Fitzpatrick, D., et al., Cancer in Northern Ireland 1993–2001: a comprehensive report, Belfast: Northern Ireland Cancer Registry; 2004.
91. Quinn, M., et al., Registrations of cancer diagnosed in 1994-1997, England & Wales in Health Statistics Quarterly 07 Autumn 2000. Office for National Statistics. p. 71-82.
92. National Statistics Online. Life Expectancy: More aged 70 and 80 than ever before. June 2004.
93. http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/risk/statistics-on-the-risk-of-developing-cancer Accessed Apr 2009
94. Office for National Statistics, Cancer Statistics registrations: registrations of cancer diagnosed in 2005, England. Series MB1 no.36., London: National Statistics.
95. ISD Online. Cancer Incidence, Mortality and Survival data.Accessed 2008
96. Welsh Cancer Intelligence and Surveillance Unit, Cancer Incidence in Wales. 2008

97. Northern Ireland Cancer Registry. Cancer Incidence and Mortality. Accessed 2008
98. http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/projections/ Accessed Apr 2009
99. http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/commoncancers/uk-cancer-incidence-statistics-for-common-cancers Accessed Apr 2009
100. Office for National Statistics, Cancer Statistics registrations: Registrations of cancer diagnosed in 2005, England
101. Welsh Cancer Intelligence and Surveillance Unit. 2008
102. ISD Online 2008. Cancer incidence and mortality data.
103. Northern Ireland Cancer Registry 2008. Cancer statistics
104. 104.http://publications.cancerresearchuk.org/downloads/Product/CSBOW6bowel.pdf Accessed Apr 2009
105. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/incidence/uk-bowel-cancer-incidence-statistics#distribution Accessed Apr 2009.
106. 106.Office for National Statistics, Mortality Statistics: Cause England & Wales, 2006. Vol. DH2 No.32. 2006: TSO.
107. 107.GRO for Scotland Registrar General's Annual Report, 2006.
108. 108.Northern Ireland Cancer Registry Cancer Mortality in Northern Ireland, 2006
109. http://www.cancerresearchuk.org/cancer-info/cancerstats/mortality/cancerdeaths/ Accessed Apr 2009
110. Office for National Statistics Mortality Statistics: Cause. England and Wales 2006 London TSO 2008
111. Northern Ireland Cancer Registry, 2008, Cancer Mortality in Northern Ireland, 2006
112. ISD Online, 2008, Cancer Mortality in Scotland, 2006
113. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/mortality/ Accessed Apr 2009
114. Quinn M, W.H., Cooper N, Rowan S, Cancer Atlas of the United Kingdom and Ireland 1991-2000 Office for National Statistics

115. NCIN: What cancer statistics are available and where can I find them. http://www.ncin.org.uk/view?rid=664 Accessed Feb 2014
116. Office for National Statistics, Cancer Statistics registrations: Registrations of cancer diagnosed in 2000, England. Series MB1 no.31. 2003, National Statistics: London.
117. ISD Online, Cancer Incidence and Mortality NHS Scotland.
118. Welsh Cancer Intelligence and Surveillance Unit, 2003
119. Northern Ireland Cancer Registry, Cancer Incidence and Mortality
120. Coleman, M., et al., Trends in socioeconomic inequalities in cancer survival in England and Wales up to 2001 in BJC, 2004.
121. Coleman, M., P. Babb, and P. Damiecki, Cancer Survival Trends in England and Wales, 1971-1995: Deprivation and NHS Region. Vol. 1999: TSO.
122. Coleman, M., P. Babb, and V. Harris, Cancer Survival in England and Wales, 1971-1995 in Health Statistics Quarterly 06 2000, Office for National Statistics.
123. Cancer Research UK, CancerStats: Survival - England and Wales, 2004.
124. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/survival/bowel-cancer-survival-statistics Accessed Apr 2009
125. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/survival/bowel-cancer-survival-statistics#age Accessed Apr 2009
126. National Cancer Intelligence Unit (NCIN) Colorectal Survival by Stage. London: ONS; 2009
127. http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/survival/bowel-cancer-survival-statistics#stage Accessed Apr 2009
128. Maddams J, Moller H and Devane C., Cancer prevalence in the UK, 2008 Thames Cancer Registry and Macmillan Cancer Support, 2008
129. Forman, D., et al., Cancer prevalence in the UK: results from the EUROPREVAL study. Ann Oncol, 2003. 14(4): p. 648-654

130. http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/prevalence/ Accessed Apr 2009.
131. http://www.nycris.nhs.uk/uploads/doc503_109_nycris_annrep2004_5.pdf Accessed 19 Feb 2014
132. http://www.nycris.nhs.uk/uploads/doc/vid_5892_nycris_annrep2008_9.pdf Accessed 19 Feb 2014-02-19
133. http://www.ons.gov.uk/ons/guide-method/index.html Accessed 19 Feb 2014
134. http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/index.html
135. http://www.ncin.org.uk/publications/reports/reports_archive
136. http://www.cancerresearchuk.org/cancer-info/cancerstats/faqs/#Why2
137. http://www.cancerresearchuk.org/cancer-info/utilities/Glossary/news-crude-rate
138. http://www.cancerresearchuk.org/cancer-info/utilities/Glossary/news-agestandardised-rate
139. http://www.nice.org.uk/nicemedia/live/11840/44860/44860.pdf Accessed Apr 2009
140. http://www.hscic.gov.uk/hes Accessed Apr 2013
141. Spencer A, Davies MP. Hospital Episode Statistics – improving the quality and value of hospital data: a national hospital e-survey of hospital consultants.BMJ Open.2012;2:e001651. http://bmjopen.bmj.com/content/2/6/e001651.full Accessed Apr 2013
142. http://www.ncbi.nlm.nih.gov/books/NBK116629/ Accessed May 2013
143. http://www.ncbi.nlm.nih.gov/books/NBK116630/ Accessed May 2013
144. http://www.ncbi.nlm.nih.gov/books/NBK116639/ Accessed May 2013
145. http://www.ncbi.nlm.nih.gov/books/NBK11662 Accessed May 2013
146. http://www.natcansat.nhs.uk/data/hes.aspx
147. http://www.natcansat.nhs.uk/data/
148. http://www.wales.nhs.uk/sitesplus/922/page/50308
149. http://www.ukacr.org/national-cancer-intelligence Accessed Mar2014

150. http://data.gov.uk/dataset/national-bowel-cancer-audit-annual-reports/resource/fd95e816-59c2-487f-b734-a20b588389f0 Accessed Apr 2009
151. http://www.ic.nhs.uk/services/national-clinical-audit-support-programme-ncasp/cancer/bowel Accessed Mar 2014
152. http://www.hscic.gov.uk/bowel Accessed Mar 2014
153. http://www.wales.nhs.uk/sites3/Documents/322/Wales_National_Bowel_Cancer_Audit_Data_Manual_-_v2_2008.pdf Accessed Apr 2009
154. http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/publicationsandstatistics/publications/publicationspolicyandguidance/DH_081006 Accessed Apr 2009
155. http://www.nycris.nhs.uk/uploads/doc502_109_nycris_3_year_development_plan.pd
156. http://www.nycris.nhs.uk/about
157. http://www.healthtalkonline.org/Overview Accessed Jul 2013
158. http://www.cancer.gov/dictionary?cdrid=445089
159. http://www.nhssurveys.org/survey/6 Accessed Mar 2013
160. http://www.ncin.org.uk/cancer_information_tools/cancer_patient_experience Accessed Mar 2014
161. Audit Commission. Data remember improving the quality of patient-based information in the NHS. Audit Commission Publications, 2002. http://www.audit-commission.gov.uk/nationalstudies/health/other/Pages/dataremember.aspx Accessed Jun 2013
162. Audit Commission. Information and data quality in the NHS, 2004. http://www.audit-commission.gov.uk/nationalstudies/health/other/Pages/informationanddataqualityinthenhs.aspx Accessed Jun 2013
163. Audit Commission. PbR data assurance framework 2008/09, 2009. http://www.audit-commission.gov.uk/nationalstudies/health/pbr/pbrdataassuranceframework200809/Pages/default.aspx Accessed Jun 2013

164. Burns EM, Rigby E, Mamidanna R. Systematic review of discharge coding accuracy. J Public Health. 2011;34:138–48.
165. http://www.ic.nhs.uk/about-us Accessed Mar 2013
166. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/215528/dh_128868.pdf
167. http://www.canceruk.net/natcansat/about.htm?index=1
168. http://www.ncin.org.uk/collecting_and_using_data/data_collection/chemotherapy.aspx
169. http://www.ncin.org.uk/collecting_and_using_data/data_collection/ncds.aspx
170. http://www.ukacr.org/future
171. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet CW for the STARD group. Towards complete and accurate reporting of studies od diagnostic accuracy: the STARD initiative. Annals of Internal Medicine. 2003;138(1):40-45
172. Guyatt GH, Tugwell, PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. CMAJ.1986;134:587-94
173. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making. 1991;11:88-94
174. Kent DL, Larson EB. Disease, level of imact, quality of research methods. Three dimensions of cliical efficacy assessment applied to magnetic resonance imaging. Invest Radiol. 1992;27:245-54
175. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostics tests and procedures. Principles and applications. Ann Intern Med. 1981;94:557-92
176. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. The selection of diagnostic tests. In: Sackett D, ed. Clinical Epidemiology. 2nd Ed Boston/Toronto/London:Little, Brown;1991.p47-57
177. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8:283-98
178. http://www.bris.ac.uk/quadas/history/ Accessed Mar 2014

179. De Bruin C, Pearse-Smith N. Building a search strategy. In: Henegan C, Perera R, Badenoch D, editors. Searching skills toolkit – finding the evidence. Oxford:Wiley-Blackwell-BMJ Books;2009.p44
180. http://www.nice.org.uk/media/633/63/The_guidelines_manual_2009_-_Appendix_G_Methodology_checklist_-_the_QUADAS_tool_for_studies_of_diagnostic_test_accuracy.pdf Accessed May 2013
181. Whiting P, Rutjes AW, Dinnes J et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technology Asssessment 2004.(8) 1-234
182. http://www.bris.ac.uk/quadas/resources/background-doc.pdf Accessed Mar 2014
183. Akiyoshi T, Oya M, Fujimoto Y, Kuroyanagi H, Ueno M, Yamaguchi T, Koyama M, Tanaka H, Matsueda K, Muto T. Comparison of preoperative whole-body positron emission tomography with MDCT in patients with primary colorectal cancer. Colorectal Disease 2009; 11:464-469
184. Arulampalam THA. FDG-PET for the pre-operative evaluation of colorectal liver metastases. Eur.J.Surg.Oncol. 2004; 30:286-291
185. Ashraf K. Colorectal carcinoma, preoperative evaluation by spiral computed tomography. Journal of the Pakistan Medical Association 2006; 56:149-153
186. Bartolozzi C, Donati F, Cioni D, Procacci C, Morana G, Chiesa A, Grazioli L, Cittadini G, Cittadini G, Giovagnoni A, Gandini G, Maass J, Lencioni R. Detection of colorectal liver metastases: a prospective multicenter trial comparing unenhanced MRI, MnDPDP-enhanced MRI, and spiral CT. Eur.Radiol. 2004; 14:14-20
187. Bhattacharjya S. B. Prospective study of contrast-enhanced computed tomography, computed tomography during arterioportography, and magnetic resonance imaging for staging colorectal liver metastases for liver resection. Br.J.Surg. 2004; 91:1361-1369
188. Cantwell CP, Setty BN, Holalkere N, Sahani DV, Fischman AJ, Blake MA. Liver Lesion Detection and Characterization in Patients

With Colorectal Cancer: A Comparison of Low Radiation Dose Non-enhanced PET/CT, Contrast-enhanced PET/CT, and Liver MRI. J.Comput.Assist.Tomogr. 2008; 32:738-744

189. Chua SC, Groves AM, Kayani I, Menezes L, Gacinovic S, Du Y, Bomanji JB, Ell PJ. The impact of F-18-FDG PET/CT in patients with liver metastases. European Journal of Nuclear Medicine and Molecular Imaging 2007; 34:1906-1914
190. Coenegrachts K, De GF, ter BL, Walgraeve N, Bipat S, Stoker J, Rigauts H. Comparison of MRI (including SS SE-EPI and SPIO-enhanced MRI) and FDG-PET/CT for the detection of colorectal liver metastases. Eur.Radiol. 2009; 19:370-379
191. Kim HJ, Kim KW, Byun JH, Won HJ, Shin YM, Kim PN, Lee MS, Lee MG. Comparison of mangafodipir trisodium- and ferucarbotran-enhanced MRI for detection and characterization of hepatic metastases in colorectal cancer patients. AJR.American journal of roentgenology. 2006; 186:1059-1066
192. Koh DM, Brown G, Riddell AM, Scurr E, Collins DJ, Allen SD, Chau I, Cunningham D, Desouza NM, Leach MO, Husband JE. Detection of colorectal hepatic metastases using MnDPDP MR imaging and diffusion-weighted imaging (DWI) alone and in combination. Eur.Radiol. 2008; 18:903-910
193. Kong G, Jackson C, Koh DM, Lewington V, Sharma B, Brown G, Cunningham D, Cook GJR. The use of F-18-FDG PET/CT in colorectal liver metastases-comparison with CT and liver MRI. European Journal of Nuclear Medicine and Molecular Imaging 2008; 35:1323-1329
194. Liu YN, Huang MX, An Q, Wei JM. The Impact of PET/CT on Therapeutic Strategy of Patients with Colorectal Cancer Metastasis. Hepatogastroenterology. 2009; 56:968-970
195. Nanashima A, Taheshita H, Sawai T, Sumida Y, Abo T, Tanaka K, Nonaka T, Sengyoku H, Hidaka S, Yasutake T, Nagayasu T. Preoperative Assessment of Liver Metastasis Originating from Colorectal Carcinoma: Is Super Paramagnetic Iron Oxide Particles-

Magnetic Resonance Imaging (SPIO-MRI) Useful for Screening? Hepatogastroenterology. 2008; 55:1750-1753

196. Orlacchio A, Schillaci O, Fusco N, Broccoli P, Maurici M, Yamgoue M, Danieli R, D'Urso S, Simonetti G. Role of PET/CT in the detection of liver metastases from colorectal cancer. Radiol.Med.(Torino). 2009; 114:571-585
197. Rappeport ED, Loft A, Berthelsen AK, von der Recke P, Larsen PN, Mogensen AM, Wettergren A, Rasmussen A, Hillingsoe J, Kirkegaard P, Thomsen C. Contrast-enhanced FDG-PET/CT vs. SPIO-enhanced MRI vs. FDG-PET vs. CT in patients with liver metastases from colorectal cancer: A prospective study with intraoperative confirmation. Acta Radiol. 2007; 48:369-378
198. Regge D, Campanella D, Anselmetti GC, Cirillo S, Gallo TM, Muratore A, Capussotti L, Galatola G, Floriani I, Aglietta M. Diagnostic accuracy of portal-phase CT and MRI with mangafodipir trisodium in detecting liver metastases from colorectal carcinoma. Clin.Radiol. 2006; 61:338-347
199. Ruers TJM. Improved selection of patients for hepatic surgery of colorectal liver metastases with 18F-FDG PET: A randomized study. J.Nucl.Med. 2009; 50:1036-1041
200. Schwartz L, Brody L, Brown K, Covey A, Tuorto S, Mazumdar M, Riedel E, Jarnagin W, Getrajdman G, Fong Y. Prospective, blinded comparison of helical CT and CT arterial portography in the assessment of hepatic metastasis from colorectal carcinoma. World J.Surg. 2006; 30:1892-1901
201. Selzner MK, Hany TF, Wildbrett P, McCormack L, Kadry Z, Clavien PA. Does the novel PET/CT imaging modality impact on the treatment of patients with metastatic colorectal cancer of the liver? Ann.Surg. 2004; 240:1027-1036
202. Truant S, Huglo D, Hebbar M, Ernst O, Steinling M, Pruvot FR. Prospective evaluation of the impact of 18Ffluoro 2 deoxy D glucose positron emission tomography of resectable colorectal liver metastases. The British journal of surgery 2005; 92:362-369

203. Vidiri A, Carpanese L, D'Annibale M, Caterino M, Cosimelli M, Zeuli M, David V, Crecco M. Evaluation of hepatic metastases from colorectal carcinoma with MR-superparamagnetic iron oxide. Journal of Experimental & Clinical Cancer Research 2004; 23:53-60
204. Wiering B, Ruers TJM, Krabbe PFM, Dekker HM, Oyen WJG. Comparison of multiphase CT, FDG-PET and intra-operative ultrasound in patients with colorectal liver metastases selected for surgery. Ann.Surg.Oncol. 2007; 14:818-826
205. Gonen M, Panageas KS, Larson SM. Statistical issues in analysis of diagnostic imaging experiments with multiple observations per patient. Radiology 2001;221:763-767
206. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
207. Wiering B, Krabbe PF, Jager GJ, Oyen WJ. The impact of fluor-18-deoxyglucose-positron emission tomography in the management of colorectal liver metastases: a systematic review and meta-analysis. Cancer.2005;104:2658-2670
208. Bipat S, Van Leeuwen MS, Comans EF, Pijl ME. Colorectal liver metastases: CT, MR imaging, and PET for diagnosis. Meta-analysis. Radiology. 2005; 237:123-131
209. http://www.med.emory.edu/EMAC/curriculum/diagnosis/reviewbias.html Accessed Jul 2013
210. http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/2_2_DIAGNOSTIC_TESTS.htm Accessed Jul 2013
211. Findlay M, Young H, Cunningham D, et al. Non invasive monitoring of tumour metabolism using fluorodeoxyglucose and positron emission tomography in colorectal cancer liver metastases: correlation with tumour response to fluorouracil. J Clin Oncol. 1996;14:700-708
212. Ruers T, Neeleman N, Jager G, et al. Value of positron emission tomography with [F-18]fluorodeoxyglucose in patients with colorectal liver metastases: a prospective study. J Clin Oncol. 2002;20:388-395

213. Young SW et al. Detection of hepatic malignancies using Mn-DPDP (manganese dipyridoxal diphosphate) hepatobiliary MRI contrast agent. Magn Reson Imaging 1990;8(3):267-276
214. http://en.wikipedia.org/wiki/Selection_bias Accessed Jul 2013
215. Strauss ES, Richardson WS, Glasziou P, Haynes RB. Evidence-Based Medicine – How to practice and teach EBM. Oxford: Elsevier-Churchill Livingstone;2005. p165-166
216. Gordon H Guyatt, Andrew D Oxman, Gunn E Vist, Regina Kunz,Yngve Falck-Ytter, Pablo Alonso-Coello, Holger J Schünemann, for the GRADE Working Group. BMJ. 2008; 336:924
217. http://www.nice.org.uk/media/633/15/The_guidelines_manual_2009_-_Appendix_C_Methodology_checklist_-_systematic_reviews_and_meta-analyses.pdf Accessed May 2013
218. http://publications.nice.org.uk/the-social-care-guidance-manual-pmg10/appendix-c-methodology-checklist-randomised-controlled-trials Accessed May 2013
219. Cairns SR, Scholefield JH, Steele RJ, Dunlop MG, Thomas HJW, Evans GD, Eaden JA, Rutter MD, Atkins WP, Saunders BP, Lucassen A, Jenkins P, Fairclough PD, Woodhouse CRJ. Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups. Gut 2010;59:666-690
220. http://www.gradeworkinggroup.org/index.htm Accessed May 2013
221. Balsham H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines:3.Rating the quality of the evidence. JClinEpi.2011;64(4):401-6
222. Jeffery M, Hickey BE, Hider PN. Follow-up strategies for patients treated for non-metastatic colorectal cancer. Cochrane Database of Systematic Reviews. 2007. Jan 24;(1): CD002200
223. Kievit J. Colorectal cancer follow-up: a reassessment of empirical evidence on effectiveness. European Journal of Surgical Oncology 2000;26:322-328.

224. Audisio RA*,* Robertson C*.* Colorectal cancer follow-up: perspectives for future studies*.* European Journal of Surgical Oncology 2000*;*26*(*4*):*329-337*.*

225. McArdle C*.* ABC of colorectal cancer. Effectiveness of follow-up*.* British Medical Journal 2000*;*321*:*1332-1335*.*

226. Gerdes H*.* Surveillance after colon cancer:is it worthwhile?*.* Gastroenterology 1990*;*99*(*6*):*1849-1851

227. NICE Technology Appraisal 176 2009 www.nice.org.uk/nicemedia/live/12216/45198/45198.pdf Accessed Aug 2013.

228. Scheithauer W, Rosen H, Kornek G-V, Sebesta C, Depisch D. Randomised comparison of combination chemotherapy plus supportive care with supportive care alone in patients with metastatic colorectal cancer. BMJ 1993; 306: 752-755.

229. Sugarbaker PH, Gianola FJ, Dwyer A, Neuman NR. A simplified plan for follow-up of patients with colon and rectal cancer supported by prospective studies of laboratory and radiologic test results. Surgery 1987;102(1):79-87.

230. Fleischer DE, Goldberg SB, Browning TH, Cooper JN, Friedman E, Goldner FH, et al. Detection and surveillance of colorectal cancer.Journal of American Medical Association 1989;261:580-585.

231. Ovaska JT, Järvinen HJ, Mecklin JP. The value of a follow-up programme after radical surgery for colorectal carcinoma. Scand J Gastroenterol. 1989 May;24(4):416–422.

232. Böhm B, Schwenk W, Hucke HP, Stock W. Does methodic long-term follow-up affect survival after curative resection of colorectal carcinoma?Dis Colon Rectum. 1993 Mar;36(3):280–286.

233. Mysliwiec PA, Brown ML, Klabunde CN, Ransohoff DF. Are physicians doing too much colonoscopy? A national survey of colorectal surveillance after polypectomy. Ann Intern Med 2004;141(4):264-71.

234. Collopy BT. The follow-up of patients after resection for large bowel cancer, May 1992. Medical Journal of Australia 1992;157(2):633-634.

235. Connor S, Bagshaw PF, Frizelle FA. Follow-up after attempted curative surgery for colorectal cancer. Postal survey of New Zealand surgeons practice. New Zealand Medical Journal 2001;114(1129):151-3.
236. Longo WE, Virgo KS, Coplin MA, Wade TP, Johnson FE. Current follow-up strategies after resection of colon cancer. Results of a survey of members of the American Society of Colon and Rectal Surgeons. Diseases of the Colon and Rectum 1994;37(6):573-583.
237. Virgo KS, Wade TP, Longo WE, Coplin MA, Vernava AM, Johnson FE. Surveillance after curative colon cancer resection: practice patterns of surgical subspecialists. Ann Surg Oncol 1995;2:472-82.
238. Tjandra JJ, Chan MKY. Follow-up after curative resection of colorectal cancer: a meta-analysis. Diseases Colon & Rectum. 2007 50(11):1783-1799
239. Wang T, Cui Y, Huang WS, Deng YH, Gong W, Li CJ, Wang JP. The role of postoperative colonoscopic surveillance after radical surgery for colorectal cancer: a prospective randomised clinical study. Gastrointestinal endoscopy 2009;69(3):609-12
240. Kjeldsen BJ, Kronborg O, Fenger C, Jorgensen OD. A prospective randomised study of follow-up after radical surgery for colorectal cancer. British Journal of Surgery 1997;84:666-669.
241. Wattchow DA, Weller DP, Esterman A, Pilotto LS, McGorm K, Hammett Z, et al. General practice vs surgical-based follow-up for patients with colon cancer: randomised controlled trial. British Journal of Cancer 2006;94:1116-1121.
242. Ohlsson B, Breland U, Ekberg H, Graffner H, Tranberg K. Follow-up after curative surgery for colorectal carcinoma. Diseases of Colon and Rectum 1995;38(6):619-626.
243. Secco GB, Fardelli R, Gianquinto D, Bonfante P, Baldi E, Ravera G, et al. Efficacy and cost of risk adapted follow-up in patients after colorectal cancer surgery: a prospective, randomized and controlled trial. European Journal of Surgical Oncology 2002;28:418-423.
244. Grossmann EM, Johnson FE, Virgo KS, Longo WE, Fossati R. Follow-up of colorectal cancer patients after resection with

curative intent—the GILDA trial. Surgical Oncology 2004;13:119-124.

245. Mueller PS, Montori VM, Bassler D, Koenig BA, Guyatt GH. Ethical issues in stopping randomised trials early because of apparent benefit. Ann Intern Med. 2007;146:878-881
246. Evans G J. Evidence based and evidence-biased medicine. Age Ageing.1995;24:461-3]
247. http://www.nice.org.uk/nicemedia/live/13597/56957/56957.pdf Accessed Mar 2014
248. Kwok H, Bisset IP, Hill GL. Preoperative staging of rectal cancer.International Journal of Colorectal Disease.2000;15(1):9-20
249. Bipat S, Glas AS, Slors FJM, Zwinderman AH, Bossuyt PMM, Stoker J. Rectal cancer: local staging and assessment of lymph node involvement with endoluminal US, CT and MR imaging –A meta-analysis.Radiology.2004;232:773-783
250. Mercury Study Group. Extramural depth of tumour invasion at thin section MR in patients with rectal cancer: results of the Mercury Study. Radiology. 2007;243(1):132-139
251. Mercury Study Group. Diagnostic accuracy of preoperative magnetic resonance imaging in predicting curative resection of rectal cancer: prospective observational study. British Medical Journal. 2006;333(7572):779
252. Kantorova I, Lipska L, Belohlavek O, Visokai V, Trubac M, Schneiderova M. Routine 18F-FDG PET preoperative staging of colorectal cancer: comparison with conventional staging and its impact on treatment decision making. Journal of Nuclear Medicine. 2003; 44(11):1784-1788
253. Llamas-Elvira JM, Rodriguez-Fernandez A, Gutierrez Sainz J, Gomez-Rio M, Bellon-Guardia M, Ramos Font C, Rebollo Aguirre AC, Cabello Garcia D, Ferron Orihuela A. Fluorine-18 fluorodeoxyglucose PET in the preoperative staging of colorectal cancer. European Journal of Nuclear Medicine and Molecular Imaging.2007;34(6):859-867v

254. Fillipone A, Ambrosini R, Fushi M, Marinelli T, Genovesi D, Bonomo L. Preoperative Tand N staging of colorectal cancer: accuracy of contrast-enhanced multi-detector row CT colonography –initial experience. Radiology. 2004;231:83-90
255. Tatli S, Mortele K, Breen E, Bleday R, Silverman S. Local staging of rectal cncer using combined pelvic phased array and endorectal coil MRI. Journal of Magnetic Resonance Imaging. 2006;23(4):534-540
256. Kim CK, Kim SH, Chun HK, Lee WY, Yun SH, Song SY, Choi D, Lim HK, Kim MJ, Lee J, Lee SJ. Preoperative staging of rectal cancer: accuracy of 3-Tesla magnetic resonance imaging. European Radiology. 2006;16(5):972-980
257. http://www.who.int/bulletin/volumes/85/10/06-039289/en
258. www.cancerresearchuk.org/cancer-info/spotcancerearly/ICBP/#What Accessed 1/11/2014
259. www.royalmarsden.nhs.uk/consultants-teams-wards/staff/Documents/gina-brown-rectal-carcinoma-staging.pdf Accessed 1 Nov 2014.
260. Dennison HJ, Dodds RM, Ntani G, Cooper Rachel, Cooper C, Aihie Sayer A, Baird J. How to get started with a systematic review in epidemiology:an introductory guide for early career researchers. Archives of Public Health.2013;71:21
261. https://www.nice.org.uk/standards-and-indicators Accessed 1/11/2014
262. Edelman LS, Guo Jia-Wen, Fraser A, Beck SL. Linking clinical research data to population databases. Nurs Res. 2013 Nov-Dec;62(6):438-444

Appendix 1

The AGREE checklist

The AGREE checklist

Domain 1: Scope and Purpose	
1. The overall objectives of the guideline are specifically described	Score 1-7
2. The health questionscovered by the guideline are described	Score 1-7
3. The population whom the guideline is to apply is described	Score 1-7
Domain 2: Stakeholder involvement	
4. The guideline development group includes individuals from all relevant professional groups	Score 1-7
5. The views and preferences of the target population (patients, public etc.) have been sought	Score 1-7
6. The target users of the guideline are clearly defined	Score 1-7
Domain 3: Rigour of development	
7. Systematic methods were used to search for the evidence	Score 1-7
8. The criteria for selecting the evidence are clearly described	Score 1-7
9. The strengths and limitations of the body of evidence are clearly described	Score 1-7
10. The methods for formulating the recommendations are clearly described	Score 1-7
11. The health benefits, side effects, and risks have been considered in formulating the recommendations.	Score 1-7
12. There is an explicit link between the recommendations and the suporting evidence	Score 1-7
13. The guideline has been externally reviewed by experts prior to its publication	Score 1-7
14. A procedure for updating the guideline is provided	Score 1-7
Domain 4: Clarity of presentation	
15. The recommendations are specific and unambiguous	Score 1-7
16.The different options for management of the health issue are clearly presented	Score 1-7
17. Key recommendations are clearly identifiable	Score 1-7
Domain 5: Applicability	
18. The guideline describes facilitators and barriers to its applications	Score 1-7
19. The guideline provides advice and / or tools on how the recommendation can be put into practice	Score 1-7
20. The potential resource implications of applying the recommendations have been considered.	Score 1-7
21. The guideline presents monitoring and/or auditing criteria	Score 1-7
Domain 6: Editorial independence	
22. The views of the funding body have not influenced the content of the guideline	Score 1-7
23. Competing interests of guideline development group members have been recorded and addressed.	Score 1-7
Reviewers overall assessment of the quality of the guideline	Score 1-7
Recommend this guideline for use	Yes modification No

Appendix 2

The QUADAS checklist

QUADAS checklist adapted for NICE

Study identification *Including author, title, reference, year of publication*	
Guideline topic:	**Review question no:**
Checklist completed by:	
	Circle one option for each question
Was the spectrum of participants representative of the patients who will receive the test in practice?	Yes No Unclear N/A
Were selection criteria clearly described?	Yes No Unclear N/A
Was the reference standard likely to classify the target condition correctly?	Yes No Unclear N/A
Was the period between performance of the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests?	Yes No Unclear N/A
Did the whole sample or a random selection of the sample receive verification using the reference standard?	Yes No Unclear N/A
Did participants receive the same reference standard regardless of the index test result?	Yes No Unclear N/A
Was the reference standard independent of the index test? (that is, the index test did not form part of the reference standard)	Yes No Unclear N/A
Was the execution of the index test described in sufficient detail to permit its replication?	Yes No Unclear N/A
Was the execution of the reference standard described in sufficient detail to permit its replication?	Yes No Unclear N/A
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes No Unclear N/A
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes No Unclear N/A
Were the same clinical data available when the test results were interpreted as would be available when the test is used in practice?	Yes No Unclear N/A
Were uninterpretable, indeterminate or intermediate test results reported?	Yes No Unclear N/A
Were withdrawals from the study explained?	Yes No Unclear N/A

[1] Adapted from: Whiting P, Rutjes AW, Dinnes J et al. (2004) Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technology Assessment 8: 1–234

Appendix 3

Evidence tables for the systematic review on diagnostic imaging for hepatic metastases from CRC

Evidence Tables for the SR on diagnostic imaging for hepatic metastases from colorectal cancer

Citation 1: Akiyoshi T, Oya M, Fujimoto Y, Kuroyanagi H, Ueno M, Yamaguchi T, Koyama M, Tanaka H, Matsueda K, Muto T. Comparison of pre-operative whole-body positron emission tomography with MDCT in patients with primary colorectal cancer. Colorectal Disease 2009; 11:464-469

Design: retrospective **Country**: Japan
Aim: to evaluate the additional value of FDG PET versus multidetector row CT (MDCT) in patients with 1° CRC

Inclusion criteria
65 patients with histologically proven colorectal cancer
patients with suspected liver or lymph node metastases
or patients with CEA >5ng/ml
or patients with low rectal cancer awaiting pre-op chemoradiotherapy to check lateral lymph node metastases

Exclusion criteria: Not specifically mentioned

Population
65 patients (36 men, 29 women)

Interventions
MDCT versus FDG PET

Outcomes
Sensitivity, specificity, PPV, NPV, Accuracy

Results 2x2 table

	Liver mets +	Liver mets -	total
CT+	22	1	23
CT -	0	42	42
total	22	43	65

Sensitivity	100% (22/22) (CI 85%-100%)
Specificity	98% (42/43) (CI 88%-100%)
PPV	96% (22/23) (CI 78%-100%)
NPV	100% (42/42) (CI 92%-100%)
Accuracy	98% (64/65) (CI 92%-100%)

	Liver mets +	Liver mets -	total
FDG PET+	20	0	20
FDG PET -	2	43	45
total	22	43	65

Sensitivity	91% (20/22) (CI 91%-99%)
Specificity	100% (43/43) (CI 92%-100%)
PPV	100% (20/20) (CI 83%-100%)
NPV	96% (43/45) (CI 85%-99%)
Accuracy	97% (63/65) (CI 89%-100%)

FDG PET failed to identify liver metastases detected by MDCT in two patients.

General comments
CT appears sufficient for detection of metastases in the liver. The strength of PET is in its ability to screen for extra-hepatic metastases and this is what leads to the change in management.

Citation 2: Arulampalam THA. FDG PET for the pre-operative evaluation of colorectal liver metastases Eur.J.Surg.Oncol. 2004; 30:286-291

Design: prospective
Country: Royal Free Hospital, UK
Aim: To assess accuracy of whole body FDG PET in the pre-operative staging of patients with CRC liver mets.

Inclusion criteria
Patients referred to a single surgeon for consideration for resection of colorectal liver metastases.
Sep 1999-May 2002
Patients had both FDG PET and spiral CT.

Exclusion criteria

Population
31 patients were studied. (median age 67, range 41-82), 15 male.
28 patients had a lesion on both PET and CT. This was considered the index lesion and only these patients were considered for assessment for resection. Follow up was for 21 months (range 5-33)
No loss to follow up.

Interventions
FDG PET
CT

Outcomes

Results
Accuracy of FDG PET and CT in detecting additional metastatic lesions in 28 patients with confirmed colorectal liver metastases.

	Liver mets +	Liver mets -	total
CT+	8	1	9
CT -	9	10	19
total	17	11	28

Sensitivity	47%
Specificity	91%
PPV	89%
NPV	53%

	Liver mets +	Liver mets -	total
FDG PET+	17	1	18
FDG PET -	0	10	10
total	17	11	28

Sensitivity	100%
Specificity	91%
PPV	94%
NPV	100%

11 patients were confirmed to have solitary liver metastases correctly demonstrated by both modalities.
10 patients were noted to have multi-focal liver metastases. All were correctly diagnosed by PET. CT was only able to identify multiple lesions in 5 patients. In 4 of these patients PET showed lesions that were not amenable to surgery. In the 5th patient laparotomy was performed. The 2nd PET lesion was not found but later identified on follow up imaging at 3 months. There was altered patient management in 12 patients (39%) (including extra-hepatic results).

General comments: FDG PET greatly adds to the decision making power of the surgical oncologist.

Citation 3: Ashraf K. Colorectal carcinoma, pre-operative evaluation by spiral computed tomography. Journal of the Pakistan Medical Association 2006; 56:149-153

Design: cross sectional prospective
Country: Pakistan
Aim: to assess the capability of spiral CT in pre-operative evaluation of colorectal carcinoma. (local spread, lymph node mets and liver mets).

Inclusion criteria
Patients with biopsy proven colorectal cancer undergoing surgery
All patients must have had the CT scan within 1 month prior to surgery

Exclusion criteria
Patients who had previous treatment for colorectal cancer or had a concurrent disease process which could result in false reading of the CT scan

Population
52 patients (32 male, 20 female,)
mean age was 58, range 22-87

Interventions
Spiral CT scan, 7mm, with gastrograffin
1 radiologist reading the images
not 'blinded' to the location of the primary tumour or the biopsy result.

Outcomes

Results

	Liver mets +	Liver mets -	total
CT+	16	2	18
CT -	2	32	34
total	18	34	52

Sensitivity	89% (CI 63.9%-98.1%)
Specificity	94% (CI 78.9%-99.0%)
PPV	89% (CI 63.9%-98.1%)
NPV	94% (CI 78.9%-99.0%)
Accuracy	92%

General comments

Citation 4: Bartolozzi C, Donati F, Cioni D, Procacci C, Morana G, Chiesa A, Grazioli L, Cittadini G, Cittadini G, Giovagnoni A, Gandini G, Maass J, Lencioni R. Detection of colorectal liver metastases: a prospective multicentre trial comparing unenhanced MRI, MnDPDP-enhanced MRI, and spiral CT. Eur.Radiol. 2004; 14:14-20

Design: prospective, multi-institutional trial
Country: Italy
Aim: to compare unenhanced MRI, MnDPDP-enhanced MRI and spiral CT in detecting hepatic CRC mets.

Inclusion criteria: Adult patient with hepatic colorectal cancer metastasis, Patient scheduled for partial hepatectomy or intra operative radio frequency thermal ablation

Exclusion criteria Pregnant or lactating woman, Severe biliary or renal insufficiency, Severe hepatic, dysfunction (Child class C), General contraindication to MRI, Inclusion in another study 7 days prior to enrolment

Population: 44 consecutive patients with colorectal hepatic metastases were examined with all 3 above modalities. 3 blinded readers interpreted the images

Interventions: unenhanced MRI, MnDPDP-enhanced MRI, spiral CT

Outcomes
primary endpoint: Sensitivity
Secondary outcome: Lesion conspicuity, quality of lesion delineation, confidence in diagnosis

Results
Per patient analysis

	Liver mets +	Liver mets -	total
CT+	22	3	25
CT -	19	0	19
total	41	3	44

Sensitivity	53.6%
Specificity	NA
PPV	88.0%
NPV	NA
Accuracy	50.0%

	Liver mets +	Liver mets -	total
MRI +	21	2	23
MRI -	21	0	21
total	42	2	44

Sensitivity	50.0%
Specificity	NA
PPV	91.3%
NPV	NA
Accuracy	47.7%

	Liver mets +	Liver mets -	total
MnDPDP MRI +	33	2	35
MnDPDP MRI -	9	0	9
total	42	2	44

Sensitivity	78.6%
Specificity	NA
PPV	94.2%
NPV	NA

Accuracy	75.0%

Per lesion analysis

	Liver mets +	Liver mets -	total
CT+	91	3?	94
CT -	37	0?	37
total	128	3?	141

Sensitivity	71%
Specificity	NA
PPV	NA
NPV	NA
Accuracy	NA

	Liver mets +	Liver mets -	total
MRI +	92	2?	
MRI -	36	0?	
total	128	2?	

Sensitivity	72%
Specificity	NA
PPV	NA
NPV	NA
Accuracy	NA

	Liver mets +	Liver mets -	total
MnDPDP MRI +	115	2?	
MnDPDP MRI -	13	0?	
total	128	2	

Sensitivity	90%
Specificity	NA
PPV	NA
NPV	NA
Accuracy	NA

Lesion size	IOUS	CT	MRI	MnDPDP MRI
<10mm	47	18(38%)	24(51%)	39(83%)
10-20mm	31	28 (90%)	24 (77%)	31(100%)
>20mm	45	45 (100%)	44 (98%)	45(100%)
All	128 (*)	91(71%)	92 (72%)	115 (90%)

* 47+31+45 = 123 not 128. This total features in all the text and tables. Either the 128 is a typographic error and all their calculations of sensitivities are based on the wrong figure or one of the sums is a typographic error. MnDPDP MRI is more sensitive than both CT (P=0.0007) and unenhanced MRI (P<0.0001) in the per lesion analysis. In the very small lesions the sensitivity difference is even more manifest. In the per patient analysis MnDPDP MRI sensitivity was higher than CT (p=0.0023) and unenhanced MRI (p=0.0013).

General comments: MnDPDP MRI is better than CT and unenhanced MRI.

Citation 5: Bhattacharjya S. B. Prospective study of contrast-enhanced computed tomography, computed tomography during arterioportography, and magnetic resonance imaging for staging colorectal liver metastases for liver resection. Br.J.Surg. 2004; 91:1361-1369

Design: prospective
Country: UK
Aim: To compare the value of contrast-enhanced CT, CT during arterioportography, and magnetic resonance imaging for staging patients with colorectal liver metastases.

Inclusion criteria
Consecutive patients between January 1996 – December 2001 with known or suspected colorectal liver metastases.

Exclusion criteria: Pulmonary metastases; intra-abdominal extra-hepatic disease (laparoscopy was performed before the laparotomy in 54 patients - suspicious nodules were biopsied, sent for frozen section, and confirmation of extra-hepatic disease contraindicated liver resection); local recurrence or metachronous primaries (all patients had colonoscopy to exclude this); medical contraindications to MRI (pacemaker, claustrophobia); medical contraindication to surgery

Population
120 patients with known or suspected colorectal liver metastases.
64 men / 56 women mean age 62 (29-74)
31 synchronous metastases – 89 metachronous metastases
85 patients had all three modalities and were finally included in the study population.
120 patients referred for consideration for resection.
120 had CT of chest, abdomen and pelvis
13 excluded after CT as either unfit for surgery or had pulmonary metastases
15 did not have an MRI due to contraindications
92 have MRI
54 of the 107 patients who had a CT and were fit for surgery proceeded to have laparoscopy (as part of another study being carried out in the unit)
7 were excluded because of peritoneal metastases
100 patients proceed to laparotomy, bimanual palpation and IOUS.
11 underwent laparotomy but no resection as they either had positive lymph nodes (4 – included in the study) or additional metastases or unfavourably positioned metastases.
89 patients went on to have liver resection

Interventions: Spiral contrast-enhanced CT (dual phase), contrast-enhanced MRI (gadolinium)
CTAP, MRI and CTAP were performed within 3 weeks of CT.
Gold standard: intraoperative ultrasound (IOUS), bimanual palpation, histology of resected specimen.

The films were reviewed by one of two consultant hepatobiliary radiologists. They were 'blinded' to the clinical history, the surgical and the pathological findings. The IOUS was performed by surgeons competent in this imaging modality and they were aware of the pre-operative findings. The pathologist that performed the histology of the resected specimens was 'blinded'.

Outcomes
Per lesion basis analysis: Sensitivity, Specificity, Positive predictive value
Per patient basis analysis

Results

The results for CTAP have been excluded from this summary as not relevant to our PICO.
It has also not been possible to extract all the information for the 2x2 tables but the summary diagnostic values have been presented.

Per lesion analysis

	Liver mets +	Liver mets -	total
CT+	176	20	196
CT -	65		

total	241		

Sensitivity	73%
Specificity	96.5%
PPV	89.8%
NPV	NA
Accuracy	NA

	Liver mets +	Liver mets -	total
GAD MRI+	154	22	176
GAD MRI -	34		
total	188		

Sensitivity	81.9%
Specificity	93.2%
PPV	87.5%
NPV	NA
Accuracy	NA

Lesion size	TOTAL	CT	GAD MRI
<10mm	42	22 of 42 (52%)	16 of 28 (57%)
>10mm	199	154 of 199 (77.4%)	138 of 150 (92%)
All	241	176 of 241 (73%)	154 of ? (86.3%)

Per patient analysis

	Liver mets +	Liver mets -	total
CT+		16	
CT -	21		
total			85?

Sensitivity	73.0%
Specificity	NA
PPV	NA
NPV	NA
Accuracy	Area under ROC curve 0.73

	Liver mets +	Liver mets -	total
GAD MRI+		18	103
GAD MRI -	16		
total	101		85?

Sensitivity	82%
Specificity	NA
PPV	NA
NPV	NA
Accuracy	Area under ROC curve 0.82

Detection of liver metastases by various imaging modalities on an individual patient basis stratified by number of lesions

Modality	Number of patients examined	Number correctly identified	Number understaged	Number overstaged
Solitary liver met				
CT	40	35	1	4

MRI	41	28	1	2
2 liver mets				
CT	28	24	3	1
MRI	22	19	1	2
3 liver mets				
CT	16	8	4	4
MRI	16	14	1	1
4 liver mets				
CT	7	4	0	3
MRI	7	3	2	2
5 liver mets				
CT	2	1	1	0
MRI	2	1	1	0
≥ 6 liver mets				
CT	7	1	6	0
MRI	7	4	3	0

Based on these results MRI is significantly superior to spiral CT (p=0.043) in staging colorectal cancer liver metastases on an individual patient basis once the number of metastases exceeds 4.
No single modality diagnosed all hepatic metastases and a multi-modal imaging approach is recommended.

General comments: The diagnostic accuracy of these modalities is similar.

Citation 6: Cantwell CP, Setty BN, Holalkere N, Sahani DV, Fischman AJ, Blake MA. Liver Lesion Detection and Characterization in Patients With Colorectal Cancer: A Comparison of Low Radiation Dose Non-enhanced PETCT, Contrast-enhanced PETCT, and Liver MRI. J.Comput.Assist.Tomogr. 2008; 32:738-744

Design: retrospective. **Country**: Boston,USA
Aim: To compare low radiation dose non-enhanced (n-e) FDG PETCT, contrast-enhanced (c-e) FDG PETCT and gadolinium-enhanced liver-specific MRI in detecting and characterising liver lesions in patients with colorectal cancer.

Inclusion criteria: Patients with colorectal cancer who had a gadolinium-enhanced MRI within 6 weeks of the PETCT scan. The follow-up diagnosis of the liver lesion must have been established either through histology of resected specimen or through imaging follow-up of at least 6 months for lesion stability or growth.
Patient should have had at least 1 but no more than 10 liver lesions
Note: previous hepatic resection and previous chemotherapy was allowed.

Exclusion criteria: More than 10 liver lesions (possibility of lesion overlap).

Population: 33 non-consecutive patients (22 men, 11 women, mean age 63 years)
retrospective review of imaging database of patients with colorectal cancer with suspected liver metastases from one institution in Boston Massachusetts from Jan 2004 to Dec 2005

Interventions: Low radiation dose non-enhanced FDG PETCT, contrast-enhanced FDG PETCT, gadolinium-enhanced liver MRI. Data was analysed by 2 radiologists. Patient demographic data was blinded as was clinical data. All data was interpreted in consensus.

Outcomes: Sensitivity, Specificity, accuracy

Results
Per lesion analysis

	Liver mets +	Liver mets -	total
Gad MRI +	98	0	98
Gad MRI -	2	10	12
total	100	10	110

Sensitivity	98%
Specificity	100%
PPV	100%
NPV	83%
Accuracy	98%

	Liver mets +	Liver mets -	total
c-e PET CT+	85	0	85
c-e PET CT -	15	10	25
total	100	10	110

Sensitivity	85%
Specificity	100%
PPV	100%
NPV	40%
Accuracy	86%

	Liver mets +	Liver mets -	total
n-e PET CT+	67	4	71
n-e PET CT -	33	6	39
total	100	10	110

Sensitivity	67%
Specificity	60%
PPV	94%

NPV	15%
Accuracy	66%

No statistically significant difference in lesion detection was found between enhanced PETCT and MRI.
Both PETCT and MRI had a higher detection rate than non-enhanced PETCT.
For lesion characterisation MRI was significantly more accurate than enhanced and non-enhanced PETCT. In turn enhanced was better than non-enhanced PETCT.

General comments
Contrast enhanced PETCT is better than unenhanced PETCT.
MRI and contrast enhanced PETCT are comparable in their detection rate
MRI is better than contrast enhanced PETCT with regard to lesion characterization.

Citation 7: Chua SC, Groves AM, Kayani I, Menezes L, Gacinovic S, Du Y, Bomanji JB, Ell PJ. The impact of F-18-FDG PETCT in patients with liver metastases. European Journal of Nuclear Medicine and Molecular Imaging 2007; 34:1906-1914

Design: retrospective
Country: UCLH London, UK
Aim: To assess the performance of PETCT versus contrast enhanced CT in the detection of colorectal liver disease.

Inclusion criteria
All patients who presented to one institution with suspected metastatic disease who underwent both PETCT and CT within 6 weeks of each other were retrospectively analysed covering a 5 year period.

Exclusion criteria

Population
131 patients
67 men, 64 women
mean age 62 (range 30-85 years)
75 had primary CRC
56 had other malignancies
patients were either pre-chemotherapy or minimum 6 weeks post-chemotherapy

Interventions
CECT (contrast enhanced CT)
FDG PETCT

Outcomes
Sensitivity, specificity, PPV, NPV
Subgroup analysis for those patients that had undergone chemotherapy (as this has the potential to alter the PETCT results)

Results
Colorectal malignancy results only

Per patient analysis

	Liver mets +	Liver mets -	total
PET CT+	63	2	65
PET CT-	4	6	10
total	67	8	75

Sensitivity	94% (CI 85%-98%)
Specificity	75% (CI 34%-96%)
PPV	97% (CI 89%-99%)
NPV	60% (CI 26%-87%)
Accuracy	NA

	Liver mets +	Liver mets -	total
c-e CT+	61	6	67
c-e CT-	6	2	8
total	67	8	75

Sensitivity	91% (CI 81%-96%)
Specificity	25% (CI 3%-65%)
PPV	91% (CI 81%-96%)
NPV	25% (CI 3%-65%)
Accuracy	NA

Subgroup analysis for patients that had and did not have chemotherapy prior to PETCT scanning.

Sensitivity - chemo	89% (CI 51%-99%)
Sensitivity – no chemo	95% (CI 85%-98%)
Specificity - chemo	100% (CI 29%-100%)
Specificity – no chemo	60% (CI 14%-94%)
PPV - chemo	100% (CI 63%-100%)
PPV – no chemo	97% (CI 87%-99%)
NPV - chemo	75% (CI 19%-99%)
NPV – no chemo	50% (CI 11%-88%)
Accuracy	NA

Chemotherapy did not affect with statistical significance the diagnostic accuracy of FDG PETCT p=0.178

General comments

FDG PETCT is more accurate than c-e CT in the detection of metastatic liver disease both from colorectal cancer and from other malignancies (only colorectal results presented here).

When the detection of extra-hepatic disease was also taken into account there was a change in management from the use of PETCT of about 25% (33 patients).

Citation 8: Coenegrachts K, De GF, ter BL, Walgraeve N, Bipat S, Stoker J, Rigauts H. Comparison of MRI (including SS SE-EPI and SPIO-enhanced MRI) and FDG PETCT for the detection of colorectal liver metastases. Eur.Radiol. 2009; 19:370-379

Design: prospective
Country: Belgium and the Netherlands
Aim: To compare prospectively the FDG PETCT and MRI in 24 consecutive patients suspected of having colorectal liver metastases.

Inclusion criteria
USS shows new non-cystic focal lesion
And / or CEA >3.4ng/ml for non-smokers, >4.3 ng/ml for smokers
ALT>41 U/L for males, >31 U/L for females
ALP >129 u/l
And /or bilirubin >1.2mg/dl
Time interval between MRI and FDG PETCT was at most 3 weeks.
Note: patients who had previously received chemotherapy for their colorectal malignancy were included, including those in whom the treatment was within a month of the FDG PETCT.

Exclusion criteria
Contraindications to MRI e.g. pacemaker, metallic implants

Population
14 men, 10 women with suspected colorectal cancer liver metastases
mean age 65.3 +/- 10.8 years
consecutive presentation between Oct 2005-Jan 2008

Interventions
FDG PETCT, MRI
All patient data was blinded. Blinded evaluations were made by 2 radiologists independently. In case of disagreement a consensus opinion was reached.
Reference standard: for lesions that were operated on - intra-operative ultrasound scan and the histology result. For lesions that were not operated on – follow-up was with repeat MRI.

Outcomes Sensitivity, Positive Predictive Value PPV

Results
Per patient analysis

	Liver mets +	Liver mets -	total
EPI MRI+	24	0	24
EPI MRI -	0	0	0
total	24	0	24

Sensitivity	100%
Specificity	NA
PPV	100%
NPV	NA
Accuracy	100%

	Liver mets +	Liver mets -	total
SPIO MRI +	24	0	24
SPIO MRI -	0	0	0
total	24	0	24

Sensitivity	100%
Specificity	NA
PPV	100%

NPV	NA
Accuracy	100%

	Liver mets +	Liver mets -	total
PET CT +	23	0	23
PET CT -	1	0	1
total	24	0	24

Sensitivity	96%
Specificity	NA
PPV	100%
NPV	NA
Accuracy	96%

Per lesion analysis
MRI and PETCT concordant in 9 patients. MRI identified more liver mets than PETCT

	Liver mets +	Liver mets -	total
EPI MRI+	77	0	77
EPI MRI -	0	0	0
total	77	0	77

Sensitivity	100%
Specificity	NA
PPV	100%
NPV	NA
Accuracy	100%

	Liver mets +	Liver mets -	total
SPIO MRI +	69	0	69
SPIO MRI -	8	0	8
total	77	0	77

Sensitivity	90%
Specificity	NA
PPV	100%
NPV	NA
Accuracy	90%

	Liver mets +	Liver mets -	total
PET CT +	47	0	47
PET CT -	30	0	30
total	77	0	77

Sensitivity	61%	NPV	NA
Specificity	NA	Accuracy	61%
PPV	100%		

Citation 9: Kim HJ, Kim KW, Byun JH, Won HJ, Shin YM, Kim PN, Lee MS, Lee MG. Comparison of mangafodipir trisodium- and ferucarbotran-enhanced MRI for detection and characterization of hepatic metastases in colorectal cancer patients. AJR.American Journal of Roentgenology. 2006; 186:1059-1066

Design: block randomisation trial
Country: South Korea
Aim: to evaluate the validity of mangafodipir trisodium- versus ferucarbotran-enhanced MRI in the detection and characterisation of hepatic lesions in colorectal cancer patients.

Inclusion criteria
Patients known to have or suspected of having hepatic metastases form colorectal cancer on the basis of prior helical CT examinations
Patients scheduled to have laparotomy for their hepatic mets or an intervention such as ablation.

Exclusion criteria :>5 hepatic metastases on CT, contraindications to MRI (pacemaker or aneurysm clip)

Population
41 patients
48 patients between June 2003 – Feb 2004 enrolled. 7 patients further excluded for multiple mets or histology confirming hepatocellular or cholangiocarcinoma.

Interventions
1.5 T MRI with either

- mangafodipir trisodium (MnDPDP) (a type of liver-specific contrast like gadolinium)
- ferucarbotran (a type of contrast used in SPIO MRI)

Outcomes

Results
PER LESION ANALYSIS

	Liver mets +	Liver mets -	total
MnDPDP MRI +	37	2	39
MnDPDP MRI -	1	0	1
total	38	2	40

Sensitivity	*97%*
Specificity	NA
PPV	*95%*
NPV	NA
Accuracy	37/40= 93%

	Liver mets +	Liver mets -	total
SPIO MRI+	31	0	31
SPIO MRI -	1	0	1
total	32	0	32

Sensitivity	*97%*
Specificity	*NA*
PPV	*100%*
NPV	*NA*
Accuracy	31/32= 97%

Citation 10: Koh DM, Brown G, Riddell AM, Scurr E, Collins DJ, Allen SD, Chau I, Cunningham D, Desouza NM, Leach MO, Husband JE. Detection of colorectal hepatic metastases using MnDPDP MR imaging and diffusion-weighted imaging (DWI) alone and in combination. Eur.Radiol. 2008; 18:903-910

Design: prospective. **Country**: Royal Marsden Oncology Hospital, UK
Aim: To compare the diagnostic accuracy of MnDPDP MRI and diffusion weighted MRI alone and combined.

Inclusion criteria
Consecutive patients with suspected colorectal liver metastatic disease
Pathologically proven adenocarcinoma of the colon or rectum
At least one liver lesion detected on CT scan or ultrasound that was diagnostic or suspicious of liver metastasis
Patients were candidates for liver resection (i.e. disease-sparing at at least two contiguous liver segments)

Exclusion criteria
Contraindication to MRI
Previous history of other malignancies.
In 5 patients no metastatic disease was diagnosed on MRI nor at follow up hence these patients were excluded from the analysis.

Population
38 consecutive patients originally referred for consideration into the study
5 patients had no evidence of metastatic disease at MRI or follow up so they were excluded.
33 patients were the final study population.
23 males, 10 females.
Mean age 57 years old (range 45-67)

Interventions
MnDPDP MRI (liver-specific contrast-enhanced MRI), DWI (diffusion weighted imaging) MRI, and the combination of both.
DWI is sensitive to the molecular diffusion of water in biological tissues and recent advancements have enabled high quality DWI images of the liver to be obtained. Breath-hold single shot echo planar diffusion-weighted (SS-EPI-DWI) MRI has been shown to be superior to SPIO liver-specific contrast-enhanced MRI.

Outcomes: ROC curve analysis with summary sensitivity and specificity.

Results
Average sensitivity and specificity from two observers reading the images of the different modalities.

	Sensitivity	Specificity
MnDPDP MRI	81.3%	93%
DWI MRI	78.3%	95%
MnDPDP + DWI MRI	92.2%	97%

	Accuracy as Area under curve from observer 1	Accuracy as Area under curve from observer 2
MnDPDP MRI	Az=0.92 (0.86-0.96)	Az=0.88 (0.82-0.93)
DWI MRI	Az=0.83 (0.76-0.89)	Az=0.90 (0.84-0.95)
MnDPDP + DWI MRI	Az 0.94 (0.89-0.98)	Az=0.96 (0.91-0.99)

There was no significant difference in the averaged sensitivities between MnDPDP and DWI modalities
For the combined MnDPDP + DWI the sensitivity was better compared with MnDPDP alone (p=0.01)
And there was a trend of improved sensitivity compared with DWI (p=0.06)
Accuracy was good but significantly improved for observer 2 who was more experienced in reading DWI images.

General comments Combination of MnDPDP and DWI improved sensitivity without loss of specificity.

Citation 11: Kong G, Jackson C, Koh DM, Lewington V, Sharma B, Brown G, Cunningham D, Cook GJR. The use of F-18-FDG PETCT in colorectal liver metastases-comparison with CT and liver MRI. European Journal of Nuclear Medicine and Molecular Imaging 2008; 35:1323-1329

Design: Retrospective **Country**: Royal Marsden, UK
Aim: to compare FDG PETCT with liver-specific contrast-enhanced MRI (Mn-DPDP) for the presence and number of liver metastases in patients with colorectal liver metastases being considered for surgery.

Inclusion criteria: Patients who had colorectal cancer and known or suspected liver metastases that were thought operable from 2004-2006 and who also had PETCT and MRI with a median time between studies <1month

Exclusion criteria: Patients with chemotherapy <3months before PETCT.

Population: 65 patients (42 men) median age 65 years with CRC and known or suspected liver metastases. Retrospective identification of patients from 2004-2006 who presented to the Royal Marsden Hospital.

Interventions PETCT, Mn-DPDP MRI. Proof of metastases in lesions operated upon came from histopathology or from MRI for those not operated on.

Outcomes: Per patient and per lesion analysis. Sensitivity, Specificity, False positives.

Results
Per patient analysis:

	Liver mets +	Liver mets -	total
MnDPDP MRI+	60	0	60
MnDPDP MRI -	1	4	5
total	61	4	65

	Mn-DPDP MRI
Sensitivity	98%
Specificity	100%

	Liver mets +	Liver mets -	total
PET CT+	60	0	60
PET CT -	1	4	5
total	61	4	65

	PET CT
Sensitivity	98%
Specificity	100%

Per lesion analysis

	Liver mets +	Liver mets -	total
MnDPDP MRI+	163	0	163
MnDPDP MRI -	2	6	8
total	165	6	171

	Mn-DPDP MRI
Sensitivity	99%
Specificity	100%

	Liver mets +	Liver mets -	total
PET CT+	155	0	155
PET CT -	10	6	16
total	165	6	171

	PETCT
Sensitivity	94%
Specificity	100%

MRI and PETCT Concordant 85% of lesions
MRI and PETCT Discordant 15% of lesions
MRI detected total 30 lesions / mean 3.8 per patient

PETCT detected 20 lesions / mean 2.5 per patient The lesions not detected by PETCT were all <1cm apart from 1 PETCT correctly identified more metastases than MRI in 1 case and confirmed metastases in an equivocal MRI lesion.
General comments: PETCT has high sensitivity and specificity for the presence of liver metastases and should be included early in the initial pre-surgical evaluation and could potentially guide the use of MRI. However MRI is superior for small liver metastases and remains a prerequisite for surgical planning.

Citation 12: Liu YN, Huang MX, An Q, Wei JM. The Impact of PETCT on Therapeutic Strategy of Patients with Colorectal Cancer Metastasis. Hepatogastroenterology. 2009; 56:968-970

Design: prospective **Country**: China
Aim: to assess the impact of PETCT on the therapeutic strategy of patients with colorectal cancer metastases.

Inclusion criteria
Patients that had suspicion of liver metastases on CT scan and CEA after resection for colorectal cancer.

Exclusion criteria

Population: 15 patients who all had contrast-enhanced CT scan and CEA and had suspicion of liver metastasis. 7 men, 8 women

Interventions: contrast-enhanced CT, PETCT

Outcomes: Sensitivity, Specificity, Change in therapeutic management

Results

	Liver mets +	Liver mets -	total
PETCT+	5 patients 9 lesions	0	5 patients 9 lesions
PETCT -	0	10 patients	10 patients
total	5 patients 9 lesions	10 patients	15 patients 9 lesions

	PETCT
Sensitivity	100%
Specificity	100%

	Liver mets +	Liver mets -	total
CT+	4 patients 6 lesions	0	4 patients 6 lesions
CT -	1 patient 3 lesions	10 patients	11 patients 3 lesions
total	5 patients 9 lesions	10 patients	15 patients 9 lesions

	PETCT
Sensitivity	80%
Specificity	100%

PETCT is statistically more sensitive than CT p=0.0009 - SIGNIFICANT

General comments: PETCT is more sensitive than contrast-enhanced CT in detecting liver metastases from colorectal cancer. Taking into account the extra-hepatic disease as well, the results of which are not presented in this review, there is a change in therapeutic strategy in 40% of patients based on the results of the PETCT.

Citation 13: Nanashima A, Taheshita H, Sawai T, Sumida Y, Abo T, Tanaka K, Nonaka T, Sengyoku H, Hidaka S, Yasutake T, Nagayasu T. Pre-operative Assessment of Liver Metastasis Originating from Colorectal Carcinoma: Is Super Paramagnetic Iron Oxide Particles-Magnetic Resonance Imaging (SPIO-MRI) Useful for Screening? Hepatogastroenterology. 2008; 55:1750-1753

Design: retrospective
Country: Japan
Aim: To examine retrospectively the accuracy of diagnosis for metastatic lesions per patient and per lesion by enhanced CT and SPIO-MRI in one institution in Japan over a 7 year period.

Inclusion criteria
Data from 47 consecutive patients with metastatic liver carcinoma who underwent hepatectomy between 2000 and June 2007 were collected retrospectively. During this period enhanced CT and SPIO-MRI were performed routinely 2 weeks before hepatic resection.
The reference standard was intra-operative ultrasound scan or palpation and histological findings in the resected specimen.

Exclusion criteria

Population
32 male, 15 female, mean age 61.4 years (24-85)
10 synchronous liver metastases (coincident with primary colorectal tumour)
35 metachronous liver metastases

Interventions Enhanced CT (dual phase multi detector), SPIO-MRI

Outcomes: Accuracy, Sensitivity, Positive predictive value, Negative predictive value

Results

- Per patient analysis:

40 of 47 patients with liver metastases were accurately diagnosed by both modalities.
Sensitivity 85% CT and SPIO-MRI
Positive predictive value 100% CT and SPIO-MRI
Negative predictive value 100% CT and SPIO-MRI
The 7 patients who were missed had small liver metastases 5-8mm.

- Per lesion analysis

Comparison of diagnosis of liver metastases between enhanced CT and SPIO-MRI in patients with liver metastases undergoing liver resection.

		Histology	Histology
		Liver mets (-)	Liver mets (+)
Enhanced CT	Liver mets (-)	15	3
Enhanced CT	Liver mets (+)	18	92
SPIO-MRI	Liver mets (-)	17	1
SPIO-MRI	Liver mets (+)	12	98

	Enhanced CT	SPIO-MRI
Sensitivity	92/110 (84%)	98/110 (89%) p=0.32
Positive predictive value PPV	92/92 (99%)	98/99 (99%)
Negative predictive value NPV	15/18 (83%)	17/18 (94%) p=0.6

Liver mets undetectable by CT in 18 lesions included 4 lesions of 5mm, 5 of 6mm, 5 of 7mm, 3 of 8mm, 1 of 9mm.
Liver mets undetectable by SPIO-MRI in 12 lesions included 4 lesions of 5mm, 4 of 6mm, 2 of 7mm, 2 of 8mm.
Conclusions
Undetectable cases had small tumours less than 8mm
In the per lesion analysis SPIO-MRI appears superior to CT but this is not statistically significant. In the per-patient analysis there was no difference between the two modalities.

Citation 14: Orlacchio A, Schillaci O, Fusco N, Broccoli P, Maurici M, Yamgoue M, Danieli R, D'Urso S, Simonetti G. Role of PETCT in the detection of liver metastases from colorectal cancer. Radiol.Med.(Torino). 2009; 114:571-585

Design: prospective. **Country**: Italy
Aim: to compare the diagnostic accuracy of FDG PET versus CT versus PETCT in the detection of liver metastases during tumour staging in patients with a diagnosis of colorectal cancer for the purposes of correct surgical planning and follow up.

Inclusion criteria / Exclusion criteria

Population: 467 patients from April 2005 to Dec 2007 with the diagnosis of CRC and suspected liver mets. 301 men, 166 women. mean age 64.4 +/-10.2 years

Interventions: CT, FDG PET, PETCT

Outcomes

Results: 426 cases (91.2%) there was concordance among the three modalities

	Liver mets +	Liver mets -	total
CT+	336	6	342
CT -	30	95	125
total	366	101	467

Sensitivity	91.07% (CI 88.02%-94.12%)
Specificity	95.42% (CI 91.84%-99.0%)
PPV	98.08% (CI 96.55%-99.6%)
NPV	80.65% (CI 74.43%-86.86%)
Accuracy	92.29% (CI 89.87%-94.71%)

	Liver mets +	Liver mets -	total
PET+	336	11	347
PET -	20	100	120
total	356	111	467

Sensitivity	94.05% (CI 91.52%-96.58%)
Specificity	91.6% (CI 86.85%-96.35%)
PPV	96.64% (CI 94.68%-98.59%)
NPV	85.71% (CI 79.92%-91.51%)
Accuracy	93.36% (CI 91.10%-95.62%)

	Liver mets +	Liver mets -	total
PETCT+	336	3	339
PETCT -	7	121	128
total	343	124	467

Sensitivity	97.92% (CI 96.39%-99.44%)
Specificity	97.71% (CI 95.15%-100%)
PPV	99.10% (CI 98.08%-100%)
NPV	94.81% (CI 91.07%-98.56%)
Accuracy	97.86% (CI 96.55%-99.17%)

There is statistically significant difference between the sensitivity, specificity and accuracy of PETCT v PET ($P<0.05$). There is also statistically significant difference between the sensitivity and accuracy of PETCT v CT ($P<0.05$). There is no difference between PET and CT.

Comments: PETCT has excellent diagnostic performance. It may modify patients treatment / have lower cost.

Citation 15: Rappeport ED, Loft A, Berthelsen AK, von der Recke P, Larsen PN, Mogensen AM, Wettergren A, Rasmussen A, Hillingsoe J, Kirkegaard P, Thomsen C. Contrast-enhanced FDG PETCT vs. SPIO-enhanced MRI vs. FDG PET vs. CT in patients with liver metastases from colorectal cancer: A prospective study with intra-operative confirmation. Acta Radiol. 2007; 48:369-378

Design: prospective
Country: Denmark
Aim: To compare PETCT with SPIO-MRI, PET, CT in the detection of liver metastases and extra-hepatic tumour from colorectal cancer.

Inclusion criteria

Exclusion criteria
Diabetes
Contraindications to MRI imaging
Timing of imaging not feasible before surgery
Extra-hepatic metastases confirmed on histology

Population
35 consecutive patients with suspected liver metastases from colorectal cancer
Patients referred between March 2004 and Nov 2005 for surgery for suspected or verified metastases
16 men, 19 women
median age 62 (range 33-74)

Interventions
PETCT
SPIO-MRI
PET
CT

Readers of the imaging studies were 'blinded' to the results of other imaging studies but were informed of the date for the primary colorectal cancer surgery.
Reference standard was intra-operative ultrasound scan and histological result of the resected specimen.

Outcomes
Sensitivity (true positives/[true positives+false negatives]
Specificity (true negatives/[true negatives+false positives]
Accuracy (true positives +true negatives) / all lesions
Positive predictive value PPV (true positives / [true positives +false positives])
Negative predictive value NPV (true negatives / [true negatives +false negatives])

Results

Per patient

	Liver mets +	Liver mets -	total
CT+	28	2	30
CT -	0	1	1
total	28	3	31

Sensitivity	100% (CI NA)
Specificity	33% (CI NA)
PPV	93% (CI NA)
NPV	100% (CI NA)
Accuracy	94% (CI NA)

	Liver mets +	Liver mets -	total
PET+	23	0	23
PET -	5	3	8
total	28	3	31

Sensitivity	82% (CI NA)
Specificity	100% (CI NA)
PPV	100% (CI NA)
NPV	38% (CI NA)
Accuracy	84% (CI NA)

	Liver mets +	Liver mets -	total
PETCT+	26	0	28
PETCT -	2	3	3
total	28	3	31

Sensitivity	93% (CI NA)
Specificity	100% (CI NA)
PPV	93% (CI NA)
NPV	100% (CI NA)
Accuracy	94%

	Liver mets +	Liver mets -	total
SPIO MRI+	28	2	30
SPIO MRI -	0	1	1
total	28	3	31

Sensitivity	100% (CI NA)
Specificity	33% (CI NA)
PPV	93% (CI NA)
NPV	100% (CI NA)
Accuracy	94

Per lesion analysis

	Liver mets +	Liver mets -	total
CT+	43	25	68
CT -	28	50	78
total	71	75	146

Sensitivity	61% (CI NA)
Specificity	67% (CI NA)
PPV	72% (CI NA)
NPV	86% (CI NA)
Accuracy	77% (CI NA)

	Liver mets +	Liver mets -	total
PET+	38	1	39
PET -	33	74	107
total	71	75	146

Sensitivity	54% (CI NA)
Specificity	99% (CI NA)
PPV	97% (CI NA)
NPV	69% (CI NA)
Accuracy	77% (CI NA)

	Liver mets +	Liver mets -	total
PETCT+	47	1	48

PETCT -	24	74	98
total	71	75	146

Sensitivity	66% (CI NA)
Specificity	99% (CI NA)
PPV	98% (CI NA)
NPV	76% (CI NA)
Accuracy	83% (CI NA)

	Liver mets +	Liver mets -	total
SPIO MRI+	58	14	72
SPIO MRI -	13	61	74
total	71	75	146

Sensitivity	82% (CI NA)
Specificity	81% (CI NA)
PPV	81% (CI NA)
NPV	82% (CI NA)
Accuracy	82% (CI NA)

Both CT and SPIO MRI were significantly more sensitive than PET alone. P<0.0001, p<0.0001 respectively and PET CT p<0.001, p<0.05 respectively.
There was no difference between SPIO MRI and CT
All modalities were more sensitive in detecting liver metastases larger than 1cm compared with liver metastases of up to 1cm. Of the 19 liver metastases that were less than 1cm in size PET diagnosed 1, PETCT 5, SPIO MRI 10 and CT 13.
There were 4 patients who had chemotherapy less than 1 month prior to PETCT. Even when these patients were excluded from the analysis CT and SPIO were significantly more sensitive than PET. (p=0.001)

General comments
PET alone was significantly less sensitive than CT and SPIO MRI in the detection of liver metastases. This conflicts with the conclusions from meta-analyses. Only some of the studies reported in the meta-analysis reported lesion by lesion sensitivity.
PETCT equalled MRI imaging in accuracy for liver metastasis detection.

Citation 16: Regge D, Campanella D, Anselmetti GC, Cirillo S, Gallo TM, Muratore A, Capussotti L, Galatola G, Floriani I, Aglietta M. Diagnostic accuracy of portal-phase CT and MRI with mangafodipir trisodium in detecting liver metastases from colorectal carcinoma. Clin.Radiol. 2006; 61:338-347

Design: prospective. **Country**: Italy
Aim: to compare the diagnostic accuracy of single section spiral CT and MRI with and without tissue-specific contrast agent MnDPDP in the detection of colorectal liver metastases.

Inclusion criteria: Consecutive patients referred to one institution undergoing surgery for primary and / or metastatic colorectal cancer.
>18 years of age. Histologically confirmed diagnosis of CRC. Surgical indication for either resection of the primary and/or liver resection of metastases according to colonoscopy and CT of chest and abdomen. Life expectancy of at least 12 weeks. Normal renal function (creatinine <1.5mg/dl).

Exclusion criteria: Pregnancy or lactation, contraindication to CT or MRI or laparoscopic surgery, CT-MRI interval > 4 weeks, CT or MRI imaging of poor quality due to movement artifact

Population
125 consecutive patients from one institution considered (Dec 2000-Mar 2003), 61 men (48.8%), Median age 64.4 (41-86). 82/125 had resection of primary. 19/82 also had synchronous metastases. 43/125 had resection of metachronous metastases. 19/125 had received neoadjuvant chemotherapy prior to inclusion in the study.

Interventions
Dual phase spiral single section CT with contrast. (Triple phase (delayed phase – done only when required by radiologist to differentiate between slowly filling haemangioma and metastasis).
MRI with and without MnDPDP contrast.
Reference Standard: IOUS combined with palpation and surgical inspection together with histopathologic reliefs (intra-operative frozen section histology when needed and histology on resected specimens).
2 radiologists assessed CT images and 2 the MRI images. Disagreement between readers was resolved by consensus re-evaluation. The readers were aware that the patient had CRC but were unaware of the result of other investigations and of the other readers. IOUS was performed by 1 of 2 radiologists and they were aware of the results of the CT and MRI.

Outcomes
Primary outcome
- sum of TP, sum of TN for all patients for CT, unenhanced MRI, MnDPDP MRI (per patient analysis)

TP = when the procedure detected the same metastases as the reference standard
TN = when the procedure correctly diagnosed no metastases.
Secondary outcome
- Sensitivity / specificity - per patient basis
- Sensitivity / PPV – per lesion basis
- The level of diagnostic confidence and inter-observer agreement

Per patient basis analysis definitions
Sensitivity = number of TP cases / number of patients with at least one metastasis.
Specificity = number of TN cases / all cases in whom the reference standard did not detect any metastases.

Results
MnDPDP MRI is more accurate than CT on a per patient basis. There is no difference between CT and MRI and only a trend of higher accuracy for MnDPDP MRI compared to unenhanced MRI.
MnDPDP MRI has a significantly higher sensitivity on a per lesion basis than both CT (OR 2.6; 95% CI 1.44, 4.92) and unenhanced MRI (OR 2.1; 95% CI 1.11, 3.84); (multiple logistic model accounting for lesion dimensions and intra-patient variability).
Kappa for inter-observer variability was 0.85 for CT, 0.77 for both enhanced and unenhanced MRI. Overall Kappa was 0.75 suggesting excellent agreement.
Diagnostic confidence levels not included in this evidence table as not a relevant outcome to PICO.
No serious side effects were reported from any of the investigations.

	CT	MRI	MnDPDP MRI	CT v MRI	CT v MnDPDP MRI	MRI v MnDPDP MRI

Per patient analysis						
Accuracy	91/125(72.8%)	98/125(78.4%)	103/125(82.4%)	p=0.071	p=0.005	P=0.059
Sensitivity	30/62(48.4%)	36/62(58.1%)	41/62(66.1%)	p=0.083	p=0.004	p=0.059
Specificity	61/63(96.8%)	62/63(98.4%)	62/63(98.4%)			
Per lesion analysis						
Sensitivity	137/191(71.7%)	143/191(74.9%)	158/191(82.7%)			
Sensitivity per lesion size						
≤ 10mm	31/65(47.7%)	35/65(53.8%)	44/65(67.7%)			
11-20mm	39/53(73.6%)	40/53(75.5%)	46/54(86.8%)			
>20mm	67/73(91.8%)	68/73(93.2%)	68/73(93.2%)			
PPV	137/163(84%)	143/149(96%)	158/165(95.8%)			

Per patient analysis

	Liver mets +	Liver mets -	total
MnDPDP MRI+	41	1	42
MnDPDP MRI -	21	62	83
total	62	63	125

Sensitivity	66.1%
Specificity	98.4%
PPV	97.6%
NPV	74.7%
Accuracy	82.4%

	Liver mets +	Liver mets -	total
MRI+	36	1	37
MRI -	26	62	88
total	62	63	125

Sensitivity	58.1%
Specificity	98.4%
PPV	97.3%
NPV	70.5%
Accuracy	78.4%

	Liver mets +	Liver mets -	total
CT+	30	2	32
CT -	32	61	93
total	62	63	125

Sensitivity	48.4%
Specificity	96.8%
PPV	94%
NPV	66%
Accuracy	72.8%

There was no difference between CT and MRI
MnDPDP MRI was more accurate and more sensitive than CT
There was a higher accuracy and sensitivity tendency for MnDPDP MRI v unenhanced MRI but not statistically significant.

Per lesion analysis

	Liver mets +	Liver mets -	total

MnDPDP MRI+	158	7	165
MnDPDP MRI -	33	67	100
total	191	74	265

Sensitivity	82.7%
Specificity	90.5%
PPV	95.8%
NPV	67.0 %
Accuracy	84.9%

	Liver mets +	Liver mets -	total
MRI+	143	6	149
MRI -	48	68	116
total	191	74	265

Sensitivity	74.9%
Specificity	91.9%
PPV	96%
NPV	58.6%
Accuracy	79.6%

	Liver mets +	Liver mets -	total
CT+	137	26	163
CT -	54	48	102
total	191	74	265

Sensitivity	71.7%
Specificity	64.9%
PPV	84%
NPV	47.1%
Accuracy	69.8%

CT and unenhanced MRI showed no difference in sensitivity in the per lesion analysis (OR 1.3, CI 0.73-2.27) The sensitivity of MnDPDP MRI was significantly higher than both CT (OR 2.6 CI 1.44-4.92), and unenhanced MRI (OR 2.1 CI 1.11-3.84)

General comments: On a per patient basis MnDPDP MRI is significantly more accurate and sensitive than CT in the detection of colorectal liver metastases. Specificity was similar. However MnDPDP MRI failed to be more accurate and sensitive than unenhanced MRI for both comparisons. There was no difference between CT and unenhanced MRI.

Citation 17: Ruers TJM. Improved selection of patients for hepatic surgery of colorectal liver metastases with 18F-FDG PET: A randomized study. J.Nucl.Med. 2009; 50:1036-1041

Design: randomised phase III multicentre trial. **Country**: the Netherlands
Aim: to investigate whether the addition of FDG PET to conventional CT-based pre-operative screening of colorectal liver metastases is beneficial and reduces the number of futile laparotomies.

Inclusion criteria: Histologically documented colorectal cancer treated by R0 resection, 1-4 suspected potentially resectable liver metastases. No evidence of extra-hepatic metastatic disease (except up to a maximum of 2 resectable lung mets on CT). No evidence of recurrent or second colorectal carcinoma on barium enema or colonoscopy. WHO performance status of 0-2. Age 18 - 75

Exclusion criteria: Previous malignancies (except in situ carcinoma of the cervix, non-melanoma cancer of the skin, or a cancer where there had been a disease-free interval of at least 10 years). Liver dysfunction (bilirubin, ALP x3 times upper limit if normal). Active infection. Poorly regulated diabetes mellitus.

Population: 150 patients with CRC liver mets selected for surgery by CT. Multicentre. May 2002 –Feb 2006.

Interventions: FDG PET and CT versus CT only

Outcomes: Primary=Number of futile laparotomies (any laparotomy that did not result in complete tumour treatment, that revealed benign disease, or that did not result in disease-free survival period > 6 months. Secondary=Disease-free survival (DFS), Overall survival (OS)

Results
Futile laparotomies

Variable	Control arm (no PET) n=75	Experimental arm (PET) n=75
No laparotomy	0	5 (7%)
Confirmed benign disease	-	2
Confirmed extra-hepatic disease	-	3
laparotomy	75 (100%)	70(93%)
Futile laparotomy	34 (45%)	21(28%)
Extra-hepatic disease at laparotomy – not resectable	6	2
Too extensive liver disease at laparotomy – not resectable	8	3
Benign disease at laparotomy	3	2
Benign disease after resection	1	1
Disease recurrence in <6 months	16	13

- A significantly higher proportion of patients underwent futile laparotomies in the control-no PET arm than in the experimental arm (45% v 28%) p=0.042
- The relative risk reduction was 38% (CI 4%-60%)
- The absolute difference of 17% means that 6 patients need to undergo PET to avoid 1 futile laparotomy.
- Futile laparotomy was not found to be associated with other prognostic factors as measured by the Fong score (p=0.539)

Survival
All patients were followed up for at least 3 years after randomization. For all patients randomized

3 year survival	Control arm (no PET)	Experimental arm (PET)
Overall survival OS	65.8%	61.3%
Disease free survival DFS	29.8%	35.5%

Both OS and DFS were not significantly different between the experimental and the control groups.

General Comments:
The introduction of PET in the pre-operative work-up of patients with suspected liver mets from colorectal cancer significantly reduces the number of futile laparotomies due to unexpected unresectable disease.

Citation 18: Schwartz L, Brody L, Brown K, Covey A, Tuorto S, Mazumdar M, Riedel E, Jarnagin W, Getrajdman G, Fong Y. Prospective, blinded comparison of helical CT and CT arterial portography in the assessment of hepatic metastasis from colorectal carcinoma. World J.Surg. 2006; 30:1892-1901

Design: prospective
Country: Memorial Sloan Kettering Cancer Centre - USA
Aim: To compare helical CT with helical CT with arterial portography aimed at detecting liver metastases from colorectal carcinoma.
Cannot obtain 2X2 table as only ROC curve presented.

Inclusion criteria

Exclusion criteria
Patients with evidence of extra-hepatic disease on imaging (37 patients)

Population
87 consecutive patients between April 1999 and April 2001 with suspected colorectal liver metastases .
all imaging done at a single institution
no evidence of extra-hepatic disease (final population analysed n=50)

Interventions
Helical CT
Helical CTAP – results not presented as not relevant to PICO

Outcomes
Sensitivity from ROC curve

Results
Only CT results are presented as they are relevant to the PICO.

	CT using cut-off 1 0-1 benign 2-3-4 malignant	CT using cut-off 2 0-1-2 benign 3-4 malignant
Sensitivity	76%	69%
Specificity	56%	82%
PPV	61%	78%
NPV	73%	75%
Accuracy	65%	76%

General comments

Citation 19: Selzner MK, Hany TF, Wildbrett P, McCormack L, Kadry Z, Clavien PA. Does the novel PETCT imaging modality impact on the treatment of patients with metastatic colorectal cancer of the liver? Ann.Surg. 2004; 240:1027-1036

Design: prospective. **Country**: Switzerland
Aim: To compare the diagnostic value of contrast enhanced CT with that of FDG PETCT in patients with metastatic colorectal cancer to the liver.

Inclusion criteria: All patients referred for consideration for liver resection between Jan 2002 and July 2003. CT and PETCT must have occurred within 2 weeks of each other.

Exclusion criteria: Synchronous metastatic lesions (metastatic liver disease coincident with the primary CRC).

Population: 76 patients, 52 men, 24 women, median age of 63 years (range 35-78), 62 patients received chemotherapy after their initial bowel resection. Median interval between chemotherapy and PETCT = 3 months (range 7 days to 15 months). Median follow up 16 months (range 6 months to 3 years).

Interventions: Contrast-enhanced CT, FDG PETCT
Follow up was at 3 and 6 months for those patients that did not proceed to surgery.
Separate CT radiologist and PET radiologist. Both 'blinded' to the results of other findings.

Outcomes
Primary outcome: Does PETCT alter the indications for surgery compared with CT.
Secondary outcome: True positives/negatives, false positives/negatives for PETCT.
The diagnostic ability of the modality in patients with a previous hepatectomy.
The influence of previous chemotherapy on the detection of tumours by PETCT.

Results
Per patient analysis

	Liver mets +	Liver mets -	total
CT+	63	3	66
CT -	3	7	10
total	66	10	76

Sensitivity	95%
Specificity	70%
PPV	95%
NPV	70%
Accuracy	92%

	Liver mets +	Liver mets -	total
PETCT+	60	1	61
PETCT -	6	9	15
total	66	10	76

Sensitivity	91%
Specificity	90%
PPV	98%
NPV	60%
Accuracy	91%

No difference between CT and PETCT with regard to specificity p=0.58

General comments
Comparable results between PETCT and CT with regard to the diagnosis of hepatic metastases.
Management is altered by PETCT but purely on the identification of extra-hepatic disease.
PETCT is also better at diagnosing recurrent liver disease in patients with prior hepatectomy.

Citation 20: Truant S, Huglo D, Hebbar M, Ernst O, Steinling M, Pruvot FR. Prospective evaluation of the impact of 18Ffluoro-2-deoxy D glucose positron emission tomography of resectable colorectal liver metastases. The British Journal of Surgery 2005; 92:362-369

Design: prospective double blind
Country: France
Aim: to assess the additional value of information provided by FDG PET over that provided by CT in patients with resectable liver metastases from colorectal cancer.

Inclusion criteria
Oct 2001-Nov 2002
Those patients whom on CT were thought to be eligible for liver resection.
If the PET was discordant with the CT this did not alter the decision to proceed to laparotomy.

Exclusion criteria

Population
All 53 patients underwent laparotomy
40 men, 13 women
mean age 63, range 44-78
27 patients presented with synchronous liver metastases. 26 had metachronous liver metastases.

Interventions
FDG PET
Helical CT, dual phase, 5mm slices, with iodinated contrast
Mean time between PET and CT was 24 days (range 0-61 days)
All PET scan performed within 2 months of laparotomy

Outcomes

Results
Per patient analysis: Unable to extract 2x2 table from descriptive statistics of the per patient analysis.

Per lesion analysis

	Liver mets +	Liver mets -	total
CT+	78	3	81
CT -	21	1	22
total	99	4	103

Sensitivity	79%
Specificity	25%
PPV	96%
NPV	5%
Accuracy	77%

	Liver mets +	Liver mets -	total
PET+	78	1	79
PET-	21	4	25
total	99	5	104

Sensitivity	79%
Specificity	80%
PPV	99%
NPV	16%
Accuracy	79%

Comments: Comparable results for PET and CT regarding liver mets. Extra lesions identified are extra-hepatic.

Citation 21: Vidiri A, Carpanese L, D'Annibale M, Caterino M, Cosimelli M, Zeuli M, David V, Crecco M. Evaluation of hepatic metastases from colorectal carcinoma with MR-superparamagnetic iron oxide. Journal of Experimental & Clinical Cancer Research 2004; 23:53-60

Design: prospective. **Country**: Italy
Aim: To compare the results obtained with SPIO-MRI and unenhanced MRI with that of spiral CT in order to select those patients suitable for liver resection.

Inclusion criteria
Patients with known colorectal neoplasm who were candidates for liver resection

Exclusion criteria
age <18
pregnancy and or lactation
hypersensitivity to administration of Dextran
stage C liver cirrhosis (Child-Pugh classification)
serious kidney insufficiency
haematological disease with splenomegaly
administration of a different contrast within 24 hours.

Population
35 patients, mean age 65, 20 men, 15 women, all potentially suitable for hepatic resection of metastatic lesions

Interventions
All patients had all the investigations.
spiral CT
SPIO-MRI (with body coil)
unenhanced MRI
All imaging was performed within 7 days
Pre- and post-op evaluation time period with a maximum of 30 days
Gold standard: IOUS combined with palpation and surgical inspection together with histopathology reliefs on resected specimens.

Outcomes
Sensitivity on a per lesion basis
Change in overall decision on a per patient basis

Results
Singularly difficult to make sense of their descriptive statistics to construct a 2x2 table.
Of the 35 patients included, 26 went to surgery and 9 did not (unresectable). Of the 9 unresectable cases 8 had chemotherapy and 1 had radiofrequency ablation.

Of patients submitted for surgery:

dimensions	No of lesions	CT	MRI	SPIO-MRI	IOUS
	48	34	32	41	48
<1cm	13	4	2	9	13
1-2cm	14	10	10	12	14
>2cm	21	20	20	20	21

3 FP on CT
2 FP on MRI
2 FP on SPIO-MRI (same as above)
5 patients were found to have unresectable disease at operation (missed by both CT and MRIs)
2 lesions considered by CT to be metastases were correctly identified by MRIs to be non-metastatic.
1 lesion identified by MRI as a metastasis and not picked up by CT at all was not a metastasis (angioma).

Of patients not submitted for surgery:

dimensions	CT	MRI	SPIO-MRI
	8	8	15

<1cm			4
1-2cm	2	2	5
>2cm	6	6	6

Per patient
In 5 cases SPIO-MRI concluded that surgery was contraindicated – the opposite to the CT conclusion (in 4 cases SPIO-MRI showed a greater number of lesions per segment, in 1 case it identified the lesion as benign and not metastatic).

Statistics
Kappa CT v MRI 0.9 good agreement
Kappa CT v SPIO-MRI 0.59 mild agreement
Kappa MRI v SPIO-MRI 0.51 mild agreement

Per patient analysis

	Liver mets +	Liver mets -	total
CT+	9+	3	
CT -	5		
total			35

Sensitivity	NA
Specificity	NA
PPV	NA
NPV	NA
Accuracy	NA

	Liver mets +	Liver mets -	total
MRI+	9+	2	
MRI -	5		
total			35

Sensitivity	NA
Specificity	NA
PPV	NA
NPV	NA
Accuracy	NA

	Liver mets +	Liver mets -	total
SPIO MRI+	9+	2	
SPIO MRI -	5		
total			35

Sensitivity	NA
Specificity	NA
PPV	NA
NPV	NA
Accuracy	NA

Per lesion analysis

	Liver mets +	Liver mets -	total
CT+	34	3	37
CT -	14		
total	48		

Sensitivity	71%
Specificity	NA

PPV	NA
NPV	NA
Accuracy	NA

	Liver mets +	Liver mets -	total
MRI+	32	2	34
MRI -	16		
total	48		

Sensitivity	66.6%
Specificity	NA
PPV	NA
NPV	NA
Accuracy	NA

	Liver mets +	Liver mets -	total
SPIO MRI+	41	2	43
SPIO MRI -	7		
total	48		

Sensitivity	85.4%
Specificity	NA
PPV	NA
NPV	NA
Accuracy	NA

McNemar test: significantly greater number lesions identified with SPIRO-MRI v MRI (p=0.008)

Citation 22: Wiering B, Ruers TJM, Krabbe PFM, Dekker HM, Oyen WJG. Comparison of multiphase CT, FDG PET and intra-operative ultrasound in patients with colorectal liver metastases selected for surgery. Ann.Surg.Oncol. 2007; 14:818-826

Design: prospective
Country: The Netherlands
Aim: to evaluate the predictive value of CT and FDG PET of the liver and extra-hepatic findings compared with findings at laparotomy and at 6 months follow-up.

Inclusion criteria
Consecutive patients between Jan 1999 and Nov 2004.
Suitable for liver resection of hepatic metastases from colorectal cancer on CT imaging.

Exclusion criteria
Presence of local recurrence on colonoscopy or colonography
No previous liver surgery
Poorly regulated diabetes

Population
131 consecutive patients thought suitable for liver resection of hepatic metastases on CT imaging.

Interventions
CT dual phase helical with intravenous contrast – iodine
PET

Outcomes
Diagnostic 2x2 tables for each modality for liver metastases, extra-hepatic, intra-abdominal and other sites.
Only liver-related results presented.

Results
Per patient analysis

	Liver mets +	Liver mets -	total
CT+	127	3	130
CT -	1	0	1
total	128	3	131

Sensitivity	99.2%
Specificity	NA
PPV	97%
NPV	NA
Accuracy	97%

	Liver mets +	Liver mets -	total
PET+	126	0	126
PET-	2	3	5
total	128	3	131

Sensitivity	98.4%
Specificity	100%
PPV	100%
NPV	60%
Accuracy	98.5%

Per lesion analysis

	Liver mets +	Liver mets -	total

CT+	257	3	260
CT -	106	0	106
total	363	3	366

Sensitivity	70.8%
Specificity	NA
PPV	98.8%
NPV	NA
Accuracy	70.2%

	Liver mets +	Liver mets -	total
PET+	260	0	260
PET-	103	3	106
total	363	3	366

Sensitivity	71.6%
Specificity	100%
PPV	100%
NPV	2.8%
Accuracy	71.8%

PET and CT both missed the majority of lesions that were smaller than 10mm. Many were only a few mm in diameter.

Detection rate of histologically-proven liver metastases

Lesion size	IOUS	CT	PET	CT and/or PET
<10mm	63	10 (16%)	10 (16%)	12 (19%)
10-20mm	172	123 (72%)	129 (75%)	142 (83%)
>20mm	128	124 (97%)	121 (95%)	125 (98%)
All	363	257 (71%)	260 (72%)	279 (77%)

Results from CT and PET may not be congruent and thus are complementary for the detection of metastases.

After 6 months follow up 42 new lesions developed in 15 patients. CT and PET had previously detected all the lesions though it had not been possible to identify them at laparotomy with palpation and IOUS.

General comments
CT and PET have similar diagnostic yield for the detection of liver metastases; both modalities are adequate on a patient basis but inadequate to detect the smallest of liver lesions. The latter finding is of limited clinical significance.

Appendix 4

NICE methodology checklist for systematic reviews and meta-analyses

NICE methodology checklist for Systematic Reviews and Meta-Analyses

Study identification:	
Guideline topic:	**Key question no:**
Checklist completed by:	

SECTION 1 : INTERNAL VALIDITY			
In a well-conducted systematic review		**In this study the criterion is (highlight the correct responses in yellow):**	
1.1	The study addresses an appropriate and clearly focused question	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.2	A description of the methodology used is included	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	The literature search is sufficiently rigorous to identify all relevant studies	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.4	Study quality is assessed and taken into account	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.5	There are enough similarities between the studies selected to make combining them reasonable	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable

SECTION 2: OVERALL ASSESSMENT OF THE STUDY		
2.1	How well was the study done to minimise bias	
2.2	If coded as + or – what is the likely direction in which bias might affect the study results	

SECTION 3: DESCRIPTION OF THE STUDY		
3.1	What types of studies are included on the review?	
3.2	How does the review help to answer your key question? Summarise the main conclusions to the review and how it relates to the relevant key question. Comment on the particular strength or weakness of the review as a source of evidence.	

Appendix 5

NICE methodology checklist for Randomised Controlled trials

NICE methodology checklist for Randomised Controlled Trials

Study identification:	
Guideline topic:	**Key question no:**
Checklist completed by:	

SECTION 1 : INTERNAL VALIDITY			
In a well-conducted RCT:		**In this study the criterion is (highlight the correct response in yellow) :**	
1.1	The study addresses an appropriate and clearly focused question	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.2	The assignment of subjects to treatment groups is randomised.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.3	An adequate concealment method is used.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.4	Subjects and investigators are kept 'blind' about treatment allocation.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.5	The treatment and control groups are similar at the start of the trial.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.6	The only difference between groups is the treatment under investigation.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.7	All relevant outcomes are measured in a standard, valid and reliable way.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.8	What percentage of the individuals or clusters recruited into each treatment arm of the study dropped out before the study was completed?	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.9	All the subjects are analysed in the groups to which they were randomly allocated (often referred to as intention-to-treat analysis).	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
1.10	Where the study is carried out at more than one site, results are comparable for all sites.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable

SECTION 2: OVERALL ASSESSMENT OF THE STUDY		
2.1	How well was the study done to minimise the risk of bias or confounding and to establish a causal relationship between exposure and effect? (select ++, + or -)	
2.2	If coded as + or – what is the likely direction in which bias might affect the study results?	
2.3	Taking into account clinical considerations, your evaluation of the methodology used and the statistical power of the study, are you certain that the overall effect is due to the study exposure?	
2.4	Are the results of this study directly applicable to the patient group targeted by this guideline?	

SECTION 3: DESCRIPTION OF THE STUDY (responses can be referred to the Evidence Table)		
3.1	How many patients are included in this study?	
3.2	What are the main characteristics of the patient population?	
3.3	What environmental or prognostic factor is being investigated?	
3.4	What comparisons are made in the study?	
3.5	How long are participants followed up in the study?	
3.6	What outcome measure(s) are used in the study?	
3.7	What size of effect is identified in the study?	
3.8	How was the study funded?	
3.9	Does this study help to answer your key question?	

Appendix 6

Evidence tables for the systematic review on the follow up of CRC

Evidence tables for the SR on CRC follow up

Citation 1: Jeffery M, Hickey BE, Hider PN. Follow-up strategies for patients treated for non-metastatic colorectal cancer. Cochrane Database of Systematic Reviews. 2007. Issue 1. 2007

Design: Systematic Review and Meta-analysis. **Country**: New Zealand
Aim: To review the available evidence concerning the benefits of intensive follow-up of colorectal cancer patients with respect to survival. Secondary endpoints included: time to diagnosis of recurrence, quality of life (QoL) and the harms and costs of surveillance and investigations.

Inclusion criteria: Only randomised controlled trials comparing different follow up strategies for patients with non-matastatic colorectal cancer (CRC) treated with curative intent were included.

Exclusion criteria: Non-randomised studies. Ongoing randomised trials (COLFOL, FACS, GILDA)

Population: Patients with non-matastatic colorectal cancer (CRC) treated with curative intent +/- adjuvant treatment. Males and females of any age with histologically proven adenocarcinoma of the colon or rectum staged as T1,2,3,4; N0,1,2; M0. Duke's stage A, B and C.

Interventions: Strategies of follow-up.
This included comparisons of
- follow-up versus no follow up
- follow-up strategies of varying intensity
- follow-up in different healthcare settings.

Follow up visits with health professionals included:
- symptom enquiry
- clinical examination
- procedures (e.g. colonoscopy)
- blood tests
- faecal analysis
- radiological examinations.

Outcomes
Primary: Overall Survival (OS)
Secondary:
- Disease specific survival
- Time to diagnosis of recurrence
- Incidence of surgery(with curative intent) for recurrence
- Interval recurrences (between planned visits)
- Quality of life
- Harms
- Cost of surveillance and investigations

Results

Eight studies were included (2141 patients in total):
- Overall survival benefit at five years exists for patients undergoing more intensive follow up
- The absolute number of recurrences was similar
- For disease free survival there is no significant survival benefit between intensive follow up and less intensive.
- There is a mortality benefit for performing more tests versus fewer tests
- There is a mortality benefit for performing liver imaging versus no liver imaging
- The weighted mean difference for the time to recurrence was significantly reduced but there was significant heterogeneity amongst the studies.
- There was significantly more curative surgical procedures in the intensively followed arm
- No useful data on quality of life, harms or cost-effectiveness were available.

Comparison	**Studies included**	**No of patients**	**Overall survival at 5 years** expressed as odds ratio (OR) and risk difference (RD) S = significant NS = not significant	**No of recurrences** expressed as odds ratio (OR) and risk difference (RD) S = significant NS = not significant
Intensive FU v minimalist FU	6 of 8	1601	OR 0.73 (CI 0.59, 0.91) **S** RD -0.06 (CI -0.11, -0.02) **S**	
	7 of 8	1938		OR 0.91 (CI 0.71, 1.1) NS RD -0.02 (CI -0.06, 0.02) NS
Clinic visit v No clinic visit	1 of 8	107	OR 0.57 (CI 0.26, 1.29) NS RD -0.12 (CI -0.3, 0.05) NS	
	2 of 8	444		OR 0.85 (CI 0.58, 1.25) NS RD -0.04 (CI -0.13, 0.05) NS
More clinic visits v Fewer clinic visits	2 of 8	804	OR 0.78 (CI 0.58, 1.05) NS RD -0.05 (CI -012 , 0.01) NS	OR 0.93 (CI 0.69, 1.26) NS RD -0.02 (CI -0.08, 0.05) NS
More tests v Fewer tests	5 of 8	1004	OR 0.64 (CI 0.49, 0.85) **S** RD -0.09 (CI -0.14, 0.03) **S**	OR 0.90 (CI 0.69, 1.16) NS RD -0.02 (CI -0.08, 0.03) NS
CEA v No CEA	1 of 8	107	OR 0.57 (CI 0.26, 1.29) NS RD -0.12 (CI -0.3, 0.05) NS	
	2 of 8	444		OR 0.85 (CI 0.58, 1.25) NS RD -0.04 (CI -0.13, 0.05) NS
Liver imaging V No liver imaging	5 of 8	1004	OR 0.64 (CI 0.49, 0.85) **S** RD -0.09 (CI -0.14, 0.03) **S**	
	6 of 8	1341		OR 0.88 (CI 0.70, 1.10) NS RD -0.03 (CI -0.08, 0.02) NS

Comparison	**Studies included**	**No of patients**	**Time to recurrence** Expressed as odds ratio (OR) in months	**Curative surgery at recurrence** Expressed as odds ratio (OR)
Intensive FU v minimalist FU	3 of 8	420	OR -6.75 (-11.06, -2.44) **S** But significant heterogeneity	
	6 of 8	1613		OR 2.41 (1.62, 3.53) **S** RD 0.06 (0.04, 0.09) **S**

Disease-free survival (DFS):
2 studies reported on DFS and their pooled result shows no significant difference in survival benefit between intensive follow up and less intensive. OR 0.92, CI (0.64, 1.31), RD-0.01 CI (-0.08, 0.05) NS.
Metachronous tumours:
7 studies reported a total of 15 metachronous tumours in the experimental arms and 9 in the control arms of the studies. 1 study reported interval tumours and noted 8 in the control and 2 in the experimental arm.

Complications:
1 study reported adverse events from follow up. 2 perforations and 2 GI bleeds from a total of 731 colonoscopies.

Quality of life:
1 study (597 patients) reported a small but significant increase in the quality of life of patients associated with more frequent follow up visits.(Kjeldsen 1997 – separate publication 1999)
A different study (203 patients) reported no difference in quality of life, anxiety, depression, and patient satisfaction in patients followed up in different settings; GP / hospital. (Wattchow 2006)

General comments

- This meta-analysis supports the general principle of follow up for patients with CRC after curative treatment. There is also a clear message that the use of liver imaging is associated with improved survival and this should be included in any follow up programme.
- However there is the limitation that the combined studies span a long time-frame during which clinical care and surgical technique have changed considerably. These factors may have an effect on survival and question the validity of applying the results of earlier studies to modern practice.
- Although there was no statistical heterogeneity amongst the studies the intensity of follow up was varied. For example the follow up intensity in the experimental arm of one study was the same as the intensity of follow up in the control arm of another study. Therefore a precise indication of frequency, type or setting of follow up cannot be extracted from the data.
- Time to recurrence was significantly less and significantly more surgical procedures were carried out in the intensively followed arms of the studies. Although this suggests that recurrences were detected earlier leading to salvage surgery that lead to the improved survival this result is subject to intervention bias. The decision for salvage surgery in these studies was made by clinicians that were not blinded. In addition there was significant heterogeneity amongst the studies that reported on time to recurrence and this result is not reliable.
- No useful data on quality of life, harms or cost-effectiveness were available.

References of Included Studies (For systematic reviews):

1. Kjeldsen BJ, Kronborg O, Fenger C, Jorgensen OD. A prospective randomised study of follow up after radical surgery for colorectal cancer. *British Journal of Surgery* 1997;84:666-669
2. Makela JT, Seppo OL, Kairaluoma MI. Five year follow up after radical surgery for colorectal cancer. *Archives of Surgery* 1995;130:1062-1067
3. Ohlsson B, Breland U, Ekberg H, Graffner H, Tranberg K. Follow up after curative surgery for colorectal carcinoma. *Diseases of colon and rectum* 1995;38(6):619-626
4. Pietra N, Sarli L, Costi R, Ouchemi C, Grattarola M, Peracchia A. Role of follow-up in management of local recurrences of colorectal cancer. *Diseases of colon and rectum* 1998;41:1127-1133
5. Rodriguez-Moranta F, Salo J, Arcusa A, Boadas J, Pinol V, Bessa X *et al*. Postoperative surveillance in patients with colorectal cancer who have undergone curative resection: A prospective, multicentre, randomised, controlled trial. *Journal of clinical oncology* 2005;24(3):1-8
6. Schoemaker D, Black R, Giles L, Toouli J. Yearly colonoscopy, liver CT and chest radiography do not influence 5-year survival of colorectal cancer patients. *Gastroenterology* 1998;114:7-14
7. Secco GB, Fardelli R, Gianquinto D, Bonfante P, Baldi E, Ravera G, *et al*. Efficacy and cost of risk adapted follow up in patients after colorectal cancer surgery: a prospective, randomised and controlled trial. *European Journal of Surgical Oncology* 2002;28:418-423
8. Wattchow DA, Weller DP, Esterman A, Pilotto LS, McGorm K, Hammett Z, *et al*. General practice versus surgical-based follow-up for patients with colon cancer: randomised controlled trial. *British Journal of Cancer* 2006;94:1116-1121

Citation 2: Tjandra JJ, Chan MKY. Follow-up after curative resection of colorectal cancer: a meta-analysis. Diseases Colon & Rectum. 2007 50(11):1783-1799

Design: Systematic review and meta-analysis. **Country**: Australia
Aim: To evaluate the impact of various follow-up intensities and strategies on the outcome of patients after curative surgery for colorectal cancer.

Inclusion criteria : All RCT that randomised at or shortly after surgery and comparing different intensities of surveillance on colorectal cancer after curative resection.

Exclusion criteria: Studies considered to have bias (studies that did not report on their randomization, inclusion and exclusion criteria,patient selection, allocation, study design)

Population: Patients with colorectal cancers that were treated surgically with curative intent. Local excision, distant metastases, inflammatory bowel disease and polyposis were excluded. Patients with co-morbidities that could not comply with follow up or in whom treatment of recurrent disease would be contraindicated were also excluded.

Interventions: Intensive follow-up strategies as defined by the different trials. The clinical assessment, the investigations as well as who delivered the follow up were to be clearly stated.

Outcomes

Mortality
Cancer-related mortality
Other cause of death
Total recurrence rate
Local recurrence rate (all and isolated)
hepatic recurrence rate (all and isolated)
lung recurrence rate
Number of Intramural recurrence
Number of Metachronous recurrences
Number of Asymptomatic recurrences
Time to recurrence
Method of detection of recurrence
Reoperation rate
Curative reoperation rate
Setting of follow up
Compliance to protocol
Complications from follow-up investigations

Results
A total of 2,923 patients were pooled from 8 RCTs

- Overall survival benefit at five years exists for patients undergoing more intensive follow up OR 0.74 (CI 0.59, 0.93) P value = 0.01
- Cancer related mortality did not show any significant difference between intensive and non-intensive follow up arms. (11.5% v 12.5%; OR 0.91; P=0.52) – grade not done.
- The number of all site recurrences was similar between the two groups. OR 0.97 (CI 0.82, 1.14)p=0.68
- However there is a significantly higher number of asymptomatic recurrences being picked up in the intensively followed up group. OR 3.42 (CI 2.17,5.41)
- There was no difference between the two groups with regard to different types of recurrence being diagnosed i.e. local, distant, intramural, metachronous, hepatic.(p>0.05)
- The weighted mean time to recurrence detection was reduced by 6 months with intensive follow up but there was significant heterogeneity among the studies pooled.
- The number of curative operations done for recurrence was significantly higher with intensive follow up. OR 2.81 (CI 1.65, 4.75)
- There was a significant survival benefit with CEA and colonoscopy. Liver USS had a significant survival benefit but CT was not found to make a significant difference to survival. Neither made a difference to recurrence detection.
- Although the number of recurrences was not significantly different more curative operations were performed for recurrence and this was the case whichever test was used for follow up.
- As far as frequency of the testing is concerned, more frequently done CEA levels was the only test associated with an improvement in overall mortality.

Comparison	**Studies included**	**No of patients**	**Overall survival at 5 years** expressed as odds ratio (OR) S = significant NS = not significant	**No of recurrences** expressed as odds ratio (OR) S = significant NS = not significant
Intensive FU v minimalist FU	8 of 8	2,923	OR 0.74 (CI 0.59, 0.93) **S** P value = 0.01	
	all site 8 of 8 Asymptomatic 6 of 8	2,923 1,679		All site OR 0.97 (CI 0.82, 1.14) NS P value=0.68 Asymptomatic OR 3.42 (CI 2.17,5.41) **S** P value<0.00001
CEA v No CEA	2 of 8	444	OR 0.57* forest plot CI **S** P value= 0.003 *****OR calculation end of table	
	2 of 8	444		OR 0.85 (CI 0.58, 1.25) NS
More CEA v Less CEA	1 of 8	207	OR 0.51* forest plot CI **S** P value=0.03	
	1 of 8	207		OR 0.83 (CI 0.61, 1.13) NS
Overall CEA V No/less CEA	3	651	OR* 0.56 forest plot CI **S** P value= 0.0002	
	3	651		0.83 (CI 0.61, 1.13) NS
Colonoscopy v no colonoscopy	4 of 8	875	OR 0.63* forest plot CI **S** P value=0.0006	
	?	538		OR 0.94 (0.39, 2.27) NS
More colonoscopy V Less colonoscopy	3of 8	1841	OR 0.96* forest plot CI NS P value 0.86	
	3 of 8	1841		OR 1.22 (0.45, 3.29) NS
Overall colonoscopy V No/less colonoscopy	7 of 8	2716	OR*0.84 forest plot CI **S** P value=0.04	
	?	432		0.87 (CI 0.37, 2.04)
USS Liver imaging V No USS liver imaging	3 of 8	702	OR 0.70* forest plot CI **S** P value=0.008	
	1 of 8	107		OR 2.77 (CI 0.51, 14.94) NS
More USS liver V Less USS liver	2 of 8	1192	OR* 0.90 forest plot CI NS P value 0.73	
	2 of 8	1192		OR 0.69 (0.44, 1.09) NS
Overall USS liver V No/less USS liver	5 of 8	1894	OR*0.84 forest plot CI NS P value 0.11	
	?	1298		0.81 (0.44, 1.5)
CT liver imaging V No CT liver imaging	6 of 8	1989	OR*0.79 forest plot CI NS P value= 0.06	
	6 of 8	1989		OR 0.99 (0.8, 1.22) NS

Comparison	**Studies included**	**No of patients**	**Time to recurrence** Expressed as odds ratio (OR) in months	**Curative surgery at recurrence** Expressed as odds ratio (OR)
Intensive FU v minimalist FU	5 of 8	1276	OR -5.91 (-8.74, -3.09) **S** But significant heterogeneity P<0.00001	
	7 of 8	707		OR 2.81 (1.65, 4.79) **S**
CEA v No CEA	2 of 8	444		OR* 2.06 forest plot CI **S** P value=0.02
More CEA v Less CEA	1 of 8	207		OR*9.86 forest plot CI **S** P value=0.0006
Overall CEA V No/less CEA	3 of 8	651		OR*2.99 forest plot CI **S** P value=0.03
Colonoscopy V No colonoscopy	4 of 8	875		OR* 1.85 forest plot CI **S** P value=0.01
More colonoscopy V Less colonoscopy	2 of 8	856		OR* 2.48 forest plot CI **S** P value=0.01
Overall colonoscopy V No/less colonoscopy	6 of 8	1731		OR*2.10 forest plot CI **S** P value=0.0006
USS Liver imaging V No USS liver imaging	3 of 8	702		OR* 1.99 forest plot CI **S** P value=0.002
More USS liver V Less USS liver	1 of 8	207		OR*9.87 forest plot CI **S** P value=0.0006
Overall USS liver V No/less USS liver	4 of 8	909		OR*2.54 forest plot CI **S** P=0.002
CT liver imaging V No CT liver imaging	5 of 8	1004		OR*2.03 forest plot CI **S** P=0.01

Complication:
1 study (Schoemaker 1998) reported complications. 4 patients (1.23%) had complications as a result of colonoscopy (2 perforations – 1 requiring laparotomy, 2 haemorrhages)

Cost:
1 study (Rodriguez 2006) included a cost analysis. Overall cost of follow up was higher with intensive follow up (300,315 euro v 188,630 euro). However, intensive follow up was more cost-effective when the respectability of recurrent disease was taken into account (16,684 euro v 18,863 euro).

General comments

- This meta-analysis supports the general principle of follow up for patients with CRC after curative treatment.
- However there is the limitation that the combined studies span a long time-frame during which clinical care and surgical technique have changed considerably. These factors may have an effect on survival and question the validity of applying the results of earlier studies to modern practice.
- Although there was no statistical heterogeneity amongst the studies the intensity of follow up was varied. For example the follow up intensity in the experimental arm of one study was the same as the intensity of follow up in the control arm of another study. Therefore a precise indication of frequency, type or setting of follow up cannot be extracted from the data.
- Time to recurrence was significantly less and significantly more surgical procedures were carried out in the intensively followed arms of the studies. Although this suggests that recurrences were detected earlier leading to salvage surgery that lead to the improved survival this result is subject to intervention bias. The decision for salvage surgery in these studies was made by clinicians that were not blinded. In addition there was significant heterogeneity amongst the studies that reported on time to recurrence and this result is not reliable.
- When looking at particular test used for follow up CEA levels and colonoscopy are the only ones that offer a significant survival benefit. The use of liver USS significantly reduced overall mortality but CT had an insignificant effect. Increasing the frequency did not improve survival or recurrence detection for any of the tests apart from CEA
- However because the contribution of individual surveillance tests varied considerably among studies and no study directly compared specific tests the optimal investigation strategy remains unclear.

References of Included Studies (For systematic reviews):

1. Kjeldsen BJ, Kronborg O, Fenger C, Jorgensen OD. A prospective randomised study of follow up after radical surgery for colorectal cancer. *British Journal of Surgery* 1997;84:666-669
2. Makela JT, Seppo OL, Kairaluoma MI. Five year follow up after radical surgery for colorectal cancer. *Archives of Surgery* 1995;130:1062-1067
3. Ohlsson B, Breland U, Ekberg H, Graffner H, Tranberg K. Follow up after curative surgery for colorectal carcinoma. *Diseases of colon and rectum* 1995;38(6):619-626
4. Pietra N, Sarli L, Costi R, Ouchemi C, Grattarola M, Peracchia A. Role of follow-up in management of local recurrences of colorectal cancer. *Diseases of colon and rectum* 1998;41:1127-1133
5. Rodriguez-Moranta F, Salo J, Arcusa A, Boadas J, Pinol V, Bessa X *et al*. Postoperative surveillance in patients with colorectal cancer who have undergone curative resection: A prospective, multicentre, randomised, controlled trial. *Journal of clinical oncology* 2005;24(3):1-8
6. Schoemaker D, Black R, Giles L, Toouli J. Yearly colonoscopy, liver CT and chest radiography do not influence 5-year survival of colorectal cancer patients. *Gastroenterology* 1998;114:7-14
7. Secco GB, Fardelli R, Gianquinto D, Bonfante P, Baldi E, Ravera G, *et al*. Efficacy and cost of risk adapted follow up in patients after colorectal cancer surgery: a prospective, randomised and controlled trial. *European Journal of Surgical Oncology* 2002;28:418-423
8. Grossmann EM, Johnson FE, Virgo KS, Longo WE, Fossati R. Follow-up of colorectal cancer patients after resection with curative intent: the GILDA trial. Surg Oncol 2004;13:119-24

Citation 3:Wang T, Cui Y, Huang WS, Deng YH, Gong W, Li CJ, Wang JP. The role of postoperative colonoscopic surveillance after radical surgery for colorectal cancer: a prospective randomised clinical study. *Gastrointestinal endoscopy* 2009;69(3):609-615.

Design: Randomised controlled trial
Country: China
Aim: To compare the efficacy of 2 different colonoscopic surveillance strategies in terms of survival and recurrence resectability.

Inclusion criteria
All patients undergoing curative resection for newly diagnosed colorectal cancer between January 1995 and March 2001. (curative resection was defined as one in which no macroscopic tumour remained at the end of the operation and histology of the specimen confirmed no tumour at the margins of resection)

Exclusion criteria
Duke's stage D, inflammatory bowel disease, familial adenomatous polyposis, hereditary non-polyposis colorectal cancer, patients over the age of 80, medical co-morbidity(making follow up difficult or 5 year survival unlikely), residence in remote area, refusal of consent.

Population
326 consecutive patients under the age of 80, undergoing curative resection for newly diagnosed colorectal cancer between Jan 1995 and Mar 2001 at a teaching hospital in China who consented to the trial, did not live in a remote area and did not have co-morbidities that made follow up difficult or 5 year survival unlikely.
7 patients were lost to follow up so there were 319 patients in the final statistical analysis.

Interventions
Colonoscopic strategy of follow up. Intensive colonoscopic surveillance (ICS) versus routine colonoscopic surveillance (RCS).

The intensive colonoscopy surveillance group (n=165)had colonoscopy at every follow-up visit i.e. 3 monthly for the first year, 6 monthly for the next 2 years and annually for the next two years.

The routine colonoscopy surveillance group (n=161) had colonoscopy performed at 6, 30 and 60 months. If colonoscopy had been preformed pre-operatively then it was not done at 6 months.

All patients were seen 3 monthly for the first year, 6 monthly for the next 2 years and annually for the next two years.
At each visit they all had
Medical history
Clinical examination
CEA levels
CXR
Liver imaging (CT or USS)

Outcomes
5 year survival rate
Numbers of post operative colorectal cancer (anastomotic recurrence and metachronous tumours)
Time to recurrence
Curative surgery for recurrence
Complications

Results
Overall survival was no different between the ICS and the RCS groups.
Patients in the ICS group had more curative operations for postoperative colorectal cancer and survived significantly longer following the detection of the postoperative colorectal cancer.
76.9% of postoperative colorectal cancers (anastomotic and metachronous) occurred within the first 2 pos-op years.
Survival

- 42 patients (26.1%) in the ICS v 50 patients (31.6%) in the RCS group died.
- No significant difference in survival seen between the two groups P=0.27
- No difference in stage or location distribution seen.

	5 year survival (%)	5 year survival (%)	P	HR (95% CI)
	ICS	RCS		
All patients	77	73	0.25	1.41 (0.92, 2.14)
Colon cancer	81	76	0.31	1.52 (0.80, 2.87)
Rectal cancer	72	70	0.49	1.32 (0.75, 2.34)
Duke's A	91	86	0.29	1.84 (0.58, 5.84)
Duke's B	76	75	0.40	1.19 (0.62,2.27)
Duke's C	63	54	0.51	1.35 (0.70, 2.59)

Postoperative colorectal cancer

- 13 patients (8.1%) in the ICS group and 18 patients (11.4%) in the RCS group had postoperative colorectal cancer detected. No significant difference between the two groups.p=0.32
- Anastomotic recurrence was diagnosed in 10 patients (6.2%) of the ICS group and 12 patients (7.6%) of the RCS group
- Metachronous tumours were diagnosed in 3 patients (1.9%) of the ICS group and 6 patients (3.8%) of the RCS group.
- 76.9% of postoperative colorectal cancers occurred within the first 2 years.

Postoperative colorectal cancer	Year 1 No / %	Year 2 No / %	Year 3 No / %	Year 4 No / %	Year 5 No / %	Later No / %
ICS (n=13)	5 (38.5%) anastomotic	5 (38.5%) 4 anastomotic 1metachronous	1 (7.7%) Anastomotic	1(7.7%) Metachronous	0	1(7.7%) Metachronous
RCS (n=18)	-	-	14 (77.8%) 10anastomotic 1metachronous	-	3 (16.7%) 2anastomotic 1metachronous	1 (5.6%) metachronous

- Significantly more patients in the ICS group were asymptomatic at the time of detection of their postoperative colorectal cancer. (OR 5.24 (1.06, 26.0) p=0.43)
- Significantly more patients in the ICS group had curative surgery for their postoperative cancer. (OR 0.12 (0.02, 0.91) p=0.31)
- Survival after recurrence was detected was significantly longer in the ICS group compared to the RCS group. (HR 2.97 (1.05,8.44) p=0.41)
- More patients that were asymptomatic were able to have curative surgery for their recurrence. 76.5% v 35.7%
- Patients with asymptomatic recurrence survived significantly longer than those who were symptomatic.p=0.005

Outcome of postoperative colorectal cancer	ICS No	ICS %	RCS No	RCS %	P value
Time to recurrence(months)	Mean 22 SD 17.6		Mean 35 SD 23.9		0.49
No of asymptomatic	10	76.9%	7	38.9%	0.04
Curative surgery for tumour recurrence	9	69.2%	6	33.3%	0.48
Survival after recurrence(months)	Mean 69.1 SD 12.3		Mean 24.4 SD 5.7		

Complications.

- 3 complications occurred in the ICS group (2 bleeds, 1 perforation)
- None in the RCS group.

General comments

- Well conducted, reasonable size RCT.
- Supports the view that intensive colonoscopic surveillance does not improve overall survival even though

meta-analysis have shown that intensive follow up in general does improve survival. • Shows that what intensive colonoscopic surveillance does achieve is earlier detection of postoperative colorectal cancer, more curative surgery for this and a longer survival following its detection. • The study also reported a large number of postoperative cancers detected in the first 2 years post op and suggests based on this finding that colonoscopy should be undertaken annually in the first two years following colorectal cancer resection.

Calculations for Tjandra 2007:

1. OR for overall survival; CEA v no CEA from forest plot. Tjandra 2007
Secco and Ohlsson (444 in total, 192+53=245 intensive arm, 145+54=199 non-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	86	
B = intensive arm no events		245-86=159
C=events in control arm	97	
D= control arm no events		199-97=102
OR = ad/bc		**OR = 0.57**

2. OR for overall survival; more CEA v less CEA from forest plot. Tjandra 2007
Pietra (207 in total, 104 intensive arm, 103 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	28	
B = intensive arm no events		104-28=76
C=events in control arm	43	
D= control arm no events		103-43=60
OR = ad/bc		**OR = 0.51**

3. OR for overall survival; CEA v less or no CEA from forest plot. Tjandra 2007
Secco, Ohlsson, Pietra (651 in total, 349 intensive arm, 302 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	114	
B = intensive arm no events		349-114=235
C=events in control arm	140	
D= control arm no events		302-140=162
OR = ad/bc		**OR = 0.56**

4. OR for curative reoperation; CEA v no CEA from forest plot. Tjandra 2007
Secco and Ohlsson (444 in total, 192+53=245 intensive arm, 145+54=199 non-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	33	
B = intensive arm no events		245-33=212
C=events in control arm	14	
D= control arm no events		199-14=185
OR = ad/bc		**OR = 2.06**

5. OR for curative reoperation; more CEA v less CEA from forest plot. Tjandra 2007
Pietra (207 in total, 104 intensive arm, 103 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	17	
B = intensive arm no events		104-17=87
C=events in control arm	2	
D= control arm no events		103-2=101
OR = ad/bc		**OR = 9.86**

6. OR for curative reoperation; CEA v less / no CEA from forest plot. Tjandra 2007
Secco, Ohlsson, Pietra (651 in total, 349 intensive arm, 302 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	50	
B = intensive arm no events		349-50=299
C=events in control arm	16	
D= control arm no events		302-16=286
OR = ad/bc		**OR = 2.99**

7. OR for overall survival; colonoscopy v no colonoscopy from forest plot. Tjandra 2007
Makela, Ohlsson, Schoemaker, Secco (875 in total, 464 intensive arm, 411 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	152	
B = intensive arm no events		464-152=312
C=events in control arm	179	
D= control arm no events		411-179=232
OR = ad/bc		**OR = 0.63**

8. OR for curative reoperation; colonoscopy v no colonoscopy from forest plot. Tjandra 2007
Makela, Ohlsson, Schoemaker, Secco (875 in total, 464 intensive arm, 411 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	44	
B = intensive arm no events		464-44=420
C=events in control arm	22	
D= control arm no events		411-22=389
OR = ad/bc		**OR = 1.85**

9. OR for overall survival; more colonoscopy v less colonoscopy from forest plot. Tjandra 2007
Kjedsen, Grossmann, Rodriguez (1841 in total, 906 intensive arm, 935 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	141	
B = intensive arm no events		906-141=765
C=events in control arm	151	
D= control arm no events		935-151=784
OR = ad/bc		**OR = 0.96**

10. OR for overall survival; colonoscopy v less / no colonoscopy from forest plot. Tjandra 2007
Kjedsen, Grossmann, Rodriguez Makela, Ohlsson, Schoemaker, Secco
(2716 in total, 1370 intensive arm, 1346 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	293	
B = intensive arm no events		1370-293=1077
C=events in control arm	330	
D= control arm no events		1346-330=1016
OR = ad/bc		**OR = 0.84**

11. OR for curative reoperation; more colonoscopy v less colonoscopy from forest plot. Tjandra 2007
Kjedsen, Rodriguez (856 in total, 417 intensive arm, 439 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	25	
B = intensive arm no events		417-25=392
C=events in control arm	11	
D= control arm no events		439-11=428
OR = ad/bc		**OR = 2.48**

12. OR for curative reoperation; colonoscopy v less /no colonoscopy from forest plot. Tjandra 2007 Makela, Ohlsson, Schoemaker, Secco, Kjedsen, Rodriguez
(1731 in total, 881 intensive arm, 850 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	69	
B = intensive arm no events		881-69=812
C=events in control arm	33	
D= control arm no events		850-33=817
OR = ad/bc		**OR = 2.10**

13. OR for overall survival; US liver v no US liver from forest plot. Tjandra 2007
Makela, Secco, Rodriguez (702 in total, 371 intensive arm, 331 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	115	
B = intensive arm no events		371-115=256
C=events in control arm	129	
D= control arm no events		331-129=202
OR = ad/bc		**OR = 0.70**

14. OR for curative reoperation; US liver v no US liver from forest plot. Tjandra 2007
Makela, Secco, Rodriguez (702 in total, 371 intensive arm, 331 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	54	
B = intensive arm no events		371-54=317
C=events in control arm	26	
D= control arm no events		331-26=305
OR = ad/bc		**OR = 1.99**

15. OR for overall survival; more USS liver v less USS liver from forest plot. Tjandra 2007
Pietra, Grossmann (1192 in total, 593 intensive arm, 599 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	60	
B = intensive arm no events		593-60=533
C=events in control arm	67	
D= control arm no events		599-67=532
OR = ad/bc		**OR = 0.90**

16. OR for overall survival; USS liver v less /no USS liver from forest plot. Tjandra 2007
Makela, Secco, Rodriguez, Pietra, Grossmann (1894 in total, 964 intensive arm, 930 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	175	
B = intensive arm no events		964-175=789
C=events in control arm	196	
D= control arm no events		930-196=734
OR = ad/bc		**OR = 0.84**

17. OR for curative reoperation; more USS liver v less USS liver from forest plot. Tjandra 2007
Pietra (207 in total, 104 intensive arm, 103 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	17	
B = intensive arm no events		104-17=87
C=events in control arm	2	
D= control arm no events		103-2=101
OR = ad/bc		**OR = 9.87**

18. OR for curative reoperation; USS liver v less / no USS liver from forest plot. Tjandra 2007
Makela, Secco, Rodriguez, Pietra (909 in total, 475 intensive arm, 434 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	71	
B = intensive arm no events		475-71=404
C=events in control arm	28	
D= control arm no events		434-28=406
OR = ad/bc		**OR = 2.54**

19. OR for overall survival; CT liver v no CT liver from forest plot. Tjandra 2007
Makela, Ohlsson, Pietra, Schoemaker, Grossmann, Rodriguez (1989 in total, 992 intensive arm, 997 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	162	
B = intensive arm no events		992-162=830
C=events in control arm	198	
D= control arm no events		997-198=799
OR = ad/bc		**OR = 0.79**

20. OR for curative reoperation; CT liver v no CT liver from forest plot. Tjandra 2007
Makela, Ohlsson, Pietra, Schoemaker, Rodriguez (1004 in total, 503 intensive arm, 501 less-intensive arm)

	Detail given	Detail calculated
A=events in intensive arm	48	
B = intensive arm no events		503-48=455
C=events in control arm	21	
D= control arm no events		501-21=480
OR = ad/bc		**OR = 2.41**

Appendix 7

GRADE tables for follow up recommendation

Question: Should intensive follow-up versus less intensive or no follow-up be recommended for non metastatic colorectal cancer?

Quality assessment							Summary of findings					Importance
							Number of patients		Effect			
Number of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Intensive follow-up	Less intensive or no follow-up	Relative (95% CI)	Absolute	Quality	
Overall survival at 5 years Jeffery et al 2007 (follow-up mean 5 years)												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	218/793 (27.5%)	274/808 (33.9%) 33.9%	OR 0.73 (0.59 to 0.91)	67 fewer per 1000 (from 21 fewer to 107 fewer) 67 fewer per 1000 (from 21 fewer to 107 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Overall survival at 5 years Tjandra 2007 (follow-up mean 5 years)												
8	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	321/1474 (21.8%)	373/1449 (25.7%) 25.7%	OR 0.74 (0.59 to 0.93)	53 fewer per 1000 (from 14 fewer to 88 fewer) 53 fewer per 1000 (from 14 fewer to 88 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Number of recurrences Jeffery 2007 (follow-up mean 5 years)												
7	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	354/985 (35.9%)	351/953 (36.8%) 36.8%	OR 0.91 (0.75 to 1.1)	22 fewer per 1000 (from 64 fewer to 22 more) 22 fewer per 1000 (from 64 fewer to 22 more)	⊕⊕⊕O MODERATE	CRITICAL
Number of recurrences (all sites) Tjandra 2007 (follow-up mean 5 years)												
8	randomised trials	serious[1]	no serious inconsistency[2]	no serious indirectness	no serious imprecision	none	429/1474 (29.1%)	417/1449 (28.8%) 28.8%	OR 0.97 (0.82 to 1.14)	6 fewer per 1000 (from 39 fewer to 28 more) 6 fewer per 1000 (from 39 fewer to 28 more)	⊕⊕⊕O MODERATE	CRITICAL

[1] the majority of studies in this comparison had unclear reporting of allocation concealment. This could introduce significant bias to the randomisation process and the results overall.
[2] heterogeneity not reported

Question continued: Should intensive follow -up versus less intensive or no follow-up be recommended for non metastatic colorectal cancer?

Number of asymptomatic recurrences Tjandra 2007 (follow-up mean 5 years)												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	162/858 (18.9%)	52/821 (6.3%) 6.3%	OR 3.42 (2.17 to 5.41)	124 more per 1000 (from 65 more to 205 more) 124 more per 1000 (from 64 more to 204 more)	⊕⊕OO LOW	CRITICAL
Time to recurrence Jeffery 2007 (follow-up mean 5 years; measured with: months from primary surgery to recurrence; Better indicated by lower values)												
3	randomised trials	serious[1]	serious[4]	no serious indirectness	no serious imprecision	none	209	211	-	MD 6.75 lower (11.06 to 2.44 lower)	⊕⊕OO LOW	CRITICAL
Time to recurrence Tjandra 2007 (follow-up mean 5 years; measured with: months from primary surgery to recurrence; Better indicated by lower values)												
5	randomised trials	serious[1]	serious[5]	no serious indirectness	no serious imprecision	none	626	650	-	MD 5.91 lower (8.74 to 3.09 lower)	⊕⊕OO LOW	CRITICAL
Curative surgery attempted for recurrence Jeffery 2007 (follow-up mean 5 years)												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	95/818 (11.6%)	40/795 (5%) 5%	OR 2.41 (1.63 to 3.54)	63 more per 1000 (from 29 more to 108 more) 63 more per 1000 (from 29 more to 107 more)	⊕⊕OO LOW	CRITICAL
Curative surgery attempted for recurrence Tjandra 2007 (follow-up mean 5 years)												
7	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	86/354 (24.3%)	35/353 (9.9%) 9.9%	OR 2.81 (1.65 to 4.79)	137 more per 1000 (from 55 more to 246 more) 137 more per 1000 (from 54 more to 246 more)	⊕⊕OO LOW	CRITICAL
Disease specific survival Jeffery 2007 (follow-up mean 5 years)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3,6]	none	73/343 (21.3%)	82/361 (22.7%) 22.7%	OR 0.92 (0.64 to 1.31)	14 fewer per 1000 (from 69 fewer to 51 more) 14 fewer per 1000 (from 69 fewer to 51 more)	⊕⊕OO LOW	CRITICAL

[1] the majority of studies in this comparison had unclear reporting of allocation concealment. This could introduce significant bias to the randomisation process and the results overall.
[2] heterogeneity not reported
[3] The total number of event is low (less than the 300 rule of thumb). This can introduce imprecision to the result.
[4] heterogeneity: p=0.00002, I squared=91%, all 3 studies favour intensive follow -up.
[5] heterogeneity: p<0.00001, I squared not given, 4 out of 5 studies favour intensive follow-up.
[6] The CI includes 1 and the lower limit is <than 0.75 and the upper limit is > 1.25

Question: Should CEA testing versus no CEA testing be recommended for non metastatic colorectal cancer follow-up ?

Quality assessment							Summary of findings					
							Number of patients		Effect		Quality	Importance
Number of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	CEA	No CEA	Relative (95% CI)	Absolute		
Overall survival at 5 years Jeffery 2007 (follow-up mean 5 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15/53 (28.3%)	22/54 (40.7%) 40.7%	OR 0.57 (0.26 to 1.29)	126 fewer per 1000 (from 256 fewer to 63 more) 126 fewer per 1000 (from 256 fewer to 63 more)	⊕⊕OO LOW	CRITICAL
Overall survival at 5 years Tjandra 2007 (follow-up mean 5 years)												
2	randomised trials	very serious[1,3]	no serious inconsistency	no serious indirectness	serious[2]	none	86/245 (35.1%)	97/199 (48.7%) 48.7%	OR 0.57 (0 to 0)[4,5]	136 fewer per 1000 (from 487 fewer to 487 fewer) 136 fewer per 1000 (from 487 fewer to 487 fewer)	⊕OOO VERY LOW	
Number of recurrences Jefferey 2007 (follow-up mean 5 years)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	118/245 (48.2%)	101/199 (50.8%) 50.8%	OR 0.85 (0.58 to 1.25)	41 fewer per 1000 (from 133 fewer to 55 more) 41 fewer per 1000 (from 133 fewer to 55 more)	⊕⊕OO LOW	CRITICAL
Number of recurrences Tjandra 2007												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	118/245 (48.2%)	101/199 (50.8%) 50.8%	OR 0.85 (0.58 to 1.25)	41 fewer per 1000 (from 133 fewer to 55 more) 41 fewer per 1000 (from 133 fewer to 55 more)	⊕⊕OO LOW	CRITICAL
Number of curative re-operations (Tjandra 2007) (follow-up mean 5 years)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	33/245 (13.5%)	14/199 (7%) 7%	OR 2.06 (0 to 0)[4,5]	65 more per 1000 (from 70 fewer to 70 fewer) 64 more per 1000 (from 70 fewer to 70 fewer)	⊕⊕OO LOW	CRITICAL

[1] unclear allocation concealment. [2] the total number of event is low (less than 300 rule of thumb)
[3] Secco trial included in the survival data though it is unclear how survival data has been extracted from this trial.
[4] for ease of comparison this OR value is my calculation based on other relevant information provided by the authors in their forest plot. Please see page of calculations in the evidence tables document for more detailed explanation of how this calculation was done. The authors reported p value and not OR.
[5] No numerical confidence interval was given but forest plot indicates that this is a statistically significant result as the diamond does not cross the line of no effect (1).

Question: Should liver imaging versus no liver imaging be recommended for non metastatic colorectal cancer follow-up?

Quality assessment							Summary of findings					Importance
							Number of patients		Effect		Quality	
Number of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Liver imaging	No liver imaging	Relative (95% CI)	Absolute		
Overall survival at 5 years Jeffery 2007 (follow-up mean 5 years)												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	130/503 (25.8%)	174/501 (34.7%) 34.7%	OR 0.64 (0.49 to 0.85)	93 fewer per 1000 (from 36 fewer to 140 fewer) 93 fewer per 1000 (from 36 fewer to 140 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Number of recurrences Jeffery 2007 (follow-up mean 5 years)												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision[2]	none	278/695 (40%)	271/646 (42%) 42%	OR 0.88 (0.7 to 1.1)	31 fewer per 1000 (from 84 fewer to 23 more) 31 fewer per 1000 (from 84 fewer to 23 more)	⊕⊕⊕O MODERATE	CRITICAL
Ultrasonography overall survival at 5 years Tjandra												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	115/371 (31%)	129/331 (39%) 39%	OR 0.7 (0 to 0)[4,5]	81 fewer per 1000 (from 390 fewer to 390 fewer) 81 fewer per 1000 (from 390 fewer to 390 fewer)	⊕⊕OO LOW	CRITICAL
CT overall survival Tjandra 2007												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	162/992 (16.3%)	198/997 (19.9%) 19.9%	OR 0.79 (0 to 0)[4,6]	35 fewer per 1000 (from 199 fewer to 199 fewer) 35 fewer per 1000 (from 199 fewer to 199 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Ultrasonography number of recurrences Tjandra 2007 (follow-up mean 5 years)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	5/52 (9.6%)	2/54 (3.7%) 3.7%	OR 2.77 (0.51 to 14.94)	59 more per 1000 (from 18 fewer to 328 more) 59 more per 1000 (from 18 fewer to 328 more)	⊕⊕OO LOW	

[1] most studies had unclear allocation concealment
[2] i do not think this comparison has significant imprecision. The total number of events is large (>300) and althought the CI includes 1 the upper limit is not >1.25
[3] the total number of event was low (less than 300 rule of thumb)

Question continued: Should liver imaging versus no liver imaging be recommended for non metastatic colorectal cancer follow-up?

CT number of recurrences (follow-up mean 5 years)												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	252/992 (25.4%)	254/997 (25.5%) 25.5%	OR 0.99 (0.8 to 1.22)	2 fewer per 1000 (from 40 fewer to 40 more) 2 fewer per 1000 (from 40 fewer to 40 more)	⊕⊕⊕O MODERATE	CRITICAL
Ultrasonography number of curative re-operations												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	54/371 (14.6%)	26/331 (7.9%) 7.9%	OR 1.99 (0 to 0)[4,5]	66 more per 1000 (from 79 fewer to 79 fewer) 67 more per 1000 (from 79 fewer to 79 fewer)	⊕⊕OO LOW	CRITICAL
CT number of curative re-operations Tjandra 2007 (follow-up mean 5 years)												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	48/503 (9.5%)	21/501 (4.2%) 4.2%	OR 2.41 (0 to 0)[4,5]	53 more per 1000 (from 42 fewer to 42 fewer) 54 more per 1000 (from 42 fewer to 42 fewer)	⊕⊕OO LOW	CRITICAL

[1] most studies had unclear allocation concealment
[2] i do not think this comparison has significant imprecision. The total number of events is large (>300) and althought the CI includes 1 the upper limit is not >1.25
[3] the total number of event was low (less than 300 rule of thumb)
[4] for ease of comparison this OR value is my calculation based on other relevant information provided by the authors in the forest plot. Please see page of calculations in the evidence tables document for more detailed explanations of how this calculation was done.
[5] No numerical confidence interval was given but forest plot indicates that this is a statistically significant result as the diamond does not cross the line of no effect(1).
[6] No numerical confidence interval was given but forest plot indicates that this is not a statistically significant result as the diamond does cross the line of no effect (1).

Question: Should intensive colonoscopic surveillance versus no colonoscopic surveillance be recommended for non metastatic colorectal cancer follow-up?

Quality assessment							Summary of findings					
							Number of patients		Effect			
Number of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Intensive colonoscopic surveillance	No colonoscopic surveillance	Relative (95% CI)	Absolute	Quality	Importance
Number of recurrences (Tjandra 2007) (follow-up 5 years)												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	11/272 (4%)	11/266 (4.1%) 4.1%	OR 0.94 (0.39 to 2.27)	2 fewer per 1000 (from 25 fewer to 48 more) 2 fewer per 1000 (from 25 fewer to 47 more)	⊕⊕OO LOW	
Overall survival at 5 years Tjandra (follow-up mean 5 years)												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	152/464 (32.8%)	179/411 (43.6%) 43.6%	OR 0.63 (0 to 0)[3,4]	108 fewer per 1000 (from 436 fewer to 436 fewer) 108 fewer per 1000 (from 436 fewer to 436 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Curative operations for recurrence (Tjandra 2007) (follow-up mean 5 years)												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	44/464 (9.5%)	22/411 (5.4%) 5.4%	OR 1.85 (0 to 0)[3,4]	41 more per 1000 (from 54 fewer to 54 fewer) 42 more per 1000 (from 54 fewer to 54 fewer)	⊕⊕OO LOW	CRITICAL

[1] most studies have unclear allocation concealment
[2] the total number of events is less than 300 (rule of thumb)
[3] No numerical confidence interval was given but forest plot indicates that this is a statistically significant result as the diamond does not cross the line of no effect (1).
[4] for ease of comparison this OR value is my calculation based on other relevant information provided by the authors in their forest plot. Please see page of calculations in the evidence tables document for more detailed explanation of how this calculation was done. The authors reported p value and not OR.

Question: Should more tests versus fewer tests be recommended for non metastatic colorectal cancer follow-up?

Quality assessment							Summary of findings					Importance
							Number of patients		Effect			
Number of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	More tests	Fewer tests	Relative (95% CI)	Absolute	Quality	
Overall survival at 5 years (follow-up mean 5 years)												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	130/503 (25.8%)	174/501 (34.7%) 89.1%	OR 0.64 (0.49 to 0.85)	93 fewer per 1000 (from 36 fewer to 140 fewer) 51 fewer per 1000 (from 17 fewer to 91 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Number of recurrences Jeffery 2007 (follow-up mean 5 years)												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision[2]	none	177/503 (35.2%)	188/501 (37.5%) 37.5%	OR 0.90 (0.69 to 1.16)	24 fewer per 1000 (from 82 fewer to 35 more) 24 fewer per 1000 (from 82 fewer to 35 more)	⊕⊕⊕O MODERATE	CRITICAL
More CEA testing versus less CEA testing for overall survival (Tjandra 2007)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	28/104 (26.9%)	43/103 (41.7%) 41.7%	OR 0.51 (0 to 0)[4,5]	150 fewer per 1000 (from 417 fewer to 417 fewer) 150 fewer per 1000 (from 417 fewer to 417 fewer)	⊕⊕OO LOW	CRITICAL
More CEA testing versus less CEA testing for recurrence (Tjandra 2007) (follow-up mean 5 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	47/104 (45.2%)	53/103 (51.5%) 51.5%	OR 0.83 (0.61 to 1.13)	47 fewer per 1000 (from 122 fewer to 30 more) 47 fewer per 1000 (from 122 fewer to 30 more)	⊕⊕OO LOW	CRITICAL
More CEA testing versus less CEA testing for curative re-operation (Tjandra 2007) (follow-up mean 5 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	17/104 (16.3%)	2/103 (1.9%) 1.9%	OR 9.86 (0 to 0)[4,5]	144 more per 1000 (from 19 fewer to 19 fewer) 141 more per 1000 (from 19 fewer to 19 fewer)	⊕⊕OO LOW	CRITICAL

[1] unclear allocation concealment
[2] I do not think this result has significant imprecision. The CI includes 1 but the number of total events is large (>300).
[3] the total number of events is very low (less than 300 rule of thumb).

Question continued: Should more tests versus fewer tests be recommended for non metastatic colorectal cancer follow-up?

More versus less colonoscopy for overall survival (Tjandra 2007) (follow-up mean 5 years)													
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	141/906 (15.6%)	151/935 (16.1%) 16.1%	OR 0.96 (0 to 0)[4,6]	5 fewer per 1000 (from 161 fewer to 161 fewer) 5 fewer per 1000 (from 161 fewer to 161 fewer)	⊕⊕OO LOW	CRITICAL
More versus less colonoscopy for overall survival (Wang 2009) (follow-up mean 5 years)													
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[3]	none	42/165 (25.5%)	50/161 (31.1%) 31.1%	HR 1.41 (0.92 to 2.14)	97 more per 1000 (from 21 fewer to 238 more) 98 more per 1000 (from 21 fewer to 238 more)	⊕⊕OO LOW	CRITICAL
More versus less colonoscopy for recurrence (Tjandra 2007) (follow-up mean 5 years)													
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	33/906 (3.6%)	32/935 (3.4%) 3.4%	OR 1.22 (0.45 to 3.29)	7 more per 1000 (from 19 fewer to 70 more) 7 more per 1000 (from 18 fewer to 70 more)	⊕⊕OO LOW	CRITICAL
More versus less colonoscopy for recurrence (anastomotic and metachronous) Wang 2009 (follow-up mean 5 years)													
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[3]	none	13/165 (7.9%)	18/161 (11.2%) 11.2%	p value 0.32 (0 to 0)	76 fewer per 1000 (from 112 fewer to 112 fewer) 76 fewer per 1000 (from 112 fewer to 112 fewer)	⊕⊕OO LOW	CRITICAL
More versus less colonoscopy for anastomotic recurrence Wang 2009 (follow-up mean 5 years)													
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[3]	none	10/13 (76.9%)	7/18 (38.9%) 38.9%	OR 5.24 (1.06 to 26)	380 more per 1000 (from 14 more to 554 more) 380 more per 1000 (from 14 more to 554 more)	⊕⊕OO LOW	CRITICAL

[1] unclear allocation concealment
[3] the total number of events is very low (less than 300 rule of thumb).
[7] there was no blinding in the study introducing high risk of performance and detection bias.

Question continued: Should more tests versus fewer tests be recommended for non metastatic colorectal cancer follow-up?

More versus less colonoscopy for time to recurrence Wang 2009 (follow-up mean 5 years; Better indicated by lower values)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	13	18	-	0.49 higher (0 to 0 higher)	⊕⊕OO LOW	CRITICAL
More versus less colonoscopy for curative operations attempted for recurrence (Tjandra 2007) (follow-up mean 5 years)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	25/417 (6%)	11/439 (2.5%) 2.5%	OR 2.48 (0 to 0)[4,5]	35 more per 1000 (from 25 fewer to 25 fewer) 35 more per 1000 (from 25 fewer to 25 fewer)	⊕⊕OO LOW	CRITICAL
More versus less colonoscopy for curative surgery attempted for recurrence (Wang 2009) (follow-up mean 5)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[3]	none	9/13 (69.2%)	8/18 (44.4%) 44.4%	OR 0.12 (0.02 to 0.91)	357 fewer per 1000 (from 23 fewer to 429 fewer) 357 fewer per 1000 (from 23 fewer to 428 fewer)	⊕⊕OO LOW	CRITICAL
More versus less colonoscopy for time of survival after recurrence(Wang 2009) (follow-up mean 5 years; Better indicated by lower values)												
1	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	13	18	-	2.97 higher (1.05 to 8.44 higher)	⊕⊕OO LOW	CRITICAL
More versus less ultrasonography for overall survival Tjandra 2007 (follow-up mean 5 years)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	60/593 (10.1%)	67/599 (11.2%) 11.2%	OR 0.90 (0 to 0)[4,6]	10 fewer per 1000 (from 112 fewer to 112 fewer) 10 fewer per 1000 (from 112 fewer to 112 fewer)	⊕⊕OO LOW	CRITICAL
More versus less ultrasonography for recurrence Tjandra 2007 (follow-up mean 5 years)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	39/593 (6.6%)	53/599 (8.8%) 8.8%	OR 0.69 (0.44 to 1.09)	26 fewer per 1000 (from 48 fewer to 7 more) 26 fewer per 1000 (from 47 fewer to 7 more)	⊕⊕OO LOW	CRITICAL
More versus less ultrasonography for curative re-operation for recurrence Tjandra 2007 (follow-up mean 5 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	17/104 (16.3%)	2/103 (1.9%) 1.9%	OR 9.87 (0 to 0)[4,5]	144 more per 1000 (from 19 fewer to 19 fewer) 141 more per 1000 (from 19 fewer to 19 fewer)	⊕⊕OO LOW	CRITICAL

[1] unclear allocation concealment
[3] the total number of events is very low (less than 300 rule of thumb).
[7] there was no blinding in the study introducing high risk of performance and detection bias.
[8] the total population size is very low (less than 400 rule of thumb for continuous outcomes)

Question: Should more tests versus fewer or no tests be recommended for follow-up for colorectal cancer?

Quality assessment							Summary of findings					Importance
							Number of patients		Effect		Quality	
Number of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	More tests	Fewer or no tests	Relative (95% CI)	Absolute		
CEA testing versus less or no CEA testing for overall recurrence Tjandra 2007 (follow-up mean 5 years)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	114/349 (32.7%)	140/302 (46.4%) 46.4%	OR 0.56 (0 to 0)[3,4]	137 fewer per 1000 (from 464 fewer to 464 fewer) 138 fewer per 1000 (from 464 fewer to 464 fewer)	⊕⊕OO LOW	CRITICAL
CEA testing versus less or no CEA testing for number of recurrences Tjandra 2007 (follow-up mean 5 years)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	165/349 (47.3%)	154/302 (51%) 51%	OR 0.83 (0.61 to 1.13)	47 fewer per 1000 (from 122 fewer to 30 more) 47 fewer per 1000 (from 122 fewer to 30 more)	⊕⊕⊕O MODERATE	CRITICAL
CEA testing versus less or no CEA testing for curative re-operation Tjandra 2007 (follow-up mean 5 years)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	50/349 (14.3%)	16/302 (5.3%) 5.3%	OR 2.99 (0 to 0)[3,4]	90 more per 1000 (from 53 fewer to 53 fewer) 90 more per 1000 (from 53 fewer to 53 fewer)	⊕⊕OO LOW	CRITICAL
Colonoscopy versus less or no colonoscopy for overall survival Tjandra 2007 (follow-up mean 5 years)												
7	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	293/1370 (21.4%)	330/1346 (24.5%) 024.5%	OR 0.84 (0 to 0)[3,4]	31 fewer per 1000 (from 245 fewer to 245 fewer) 31 fewer per 1000 (from 245 fewer to 245 fewer)	⊕⊕⊕O MODERATE	CRITICAL

[1] Allocation concealment unclear in the majority of studies.
[2] total number of events is less than 300 (rule of thumb)
[3] for ease of comparison the OR value is my own calculation based on other relevant information provided by the authors in their forest plot. Please see page of caluclations in the evidence tables document for more detailed explanation of how this calculation was made. The authors reported p-value not OR.
[4] No numerical confidence interval was given but forest plot indicates that this is a statistically significant result as the diamond does not cross the line of no effect (1).
[5] No numerical confidence interval was given but forest plot indicates that this is not a statistically significant result as the diamond does cross the line of no effect (1).

Question continued: Should more tests versus fewer or no tests be recommended for follow-up for colorectal cancer?

Colonoscopy versus less or no colonoscopy for number of recurrences Tjandra 2007 (follow-up mean 5 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	44/1178 (3.7%)	43/1201 (3.6%) 3.6%	OR 1.09 (0.6 to 1.98)	3 more per 1000 (from 14 fewer to 33 more) 3 more per 1000 (from 14 fewer to 33 more)	⊕⊕OO LOW	CRITICAL
Colonoscopy versus less or no colonoscopy for curative re-operation Tjandra 2007 (follow-up mean 5 years)												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	69/881 (7.8%)	33/850 (3.9%) 3.9%	OR 2.10 (0 to 0)[3,4]	39 more per 1000 (from 39 fewer to 39 fewer) 40 more per 1000 (from 39 fewer to 39 fewer)	⊕⊕OO LOW	CRITICAL
Ultrasonography versus less or no ultrasonography of the liver for overall survival Tjandra 2007 (follow-up mean 5 years)												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	175/964 (18.2%)	196/930 (21.1%) 21.1%	OR 0.84 (0 to 0)[3,5]	28 fewer per 1000 (from 211 fewer to 211 fewer) 28 fewer per 1000 (from 211 fewer to 211 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Ultrasonography versus less or no ultrasonography for number of recurrences Tjandra 2007 (follow-up mean 5 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	44/645 (6.8%)	55/653 (8.4%) 8.4%	OR 0.81 (0.44 to 1.5)	15 fewer per 1000 (from 45 fewer to 37 more) 15 fewer per 1000 (from 45 fewer to 37 more)	⊕⊕OO LOW	CRITICAL
Ultrasonography versus less or no ultrasonography for curative re-operations Tjandra 2007												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	71/475 (14.9%)	28/434 (6.5%) 6.5%	OR 2.54 (0 to 0)[3,4]	85 more per 1000 (from 65 fewer to 65 fewer) 85 more per 1000 (from 65 fewer to 65 fewer)	⊕⊕OO LOW	CRITICAL

[1] Allocation concealment unclear in the majority of studies.
[2] total number of events is less than 300 (rule of thumb)
[3] for ease of comparison the OR value is my own calculation based on other relevant information provided by the authors in their forest plot. Please see page of caluclations in the evidence tables document for more detailed explanation of how this calculation was made. The authors reported p-value not OR.
[4] No numerical confidence interval was given but forest plot indicates that this is a statistically significant result as the diamond does not cross the line of no effect (1).
[5] No numerical confidence interval was given but forest plot indicates that this is not a statistically significant result as the diamond does cross the line of no effect (1).

Question: Should follow -up clinic visits versus no follow -up clinic visit be recommended in non metastatic colorectal cancer suveillance?

Quality assessment							Summary of findings					
							Number of patients		Effect		Quality	Importance
Number of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Clinic visit	No clinic visit	Relative (95% CI)	Absolute		
Overall survival at 5 years Jeffery 2007 (follow-up mean 5 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2,3]	none	15/53 (28.3%)	22/54 (40.7%) 40.7%	OR 0.57 (0.26 to 1.29)	126 fewer per 1000 (from 256 fewer to 63 more) 126 fewer per 1000 (from 256 fewer to 63 more)	⊕⊕OO LOW	CRITICAL
Number of recurrences Jeffery 2007 (follow-up mean 5 years)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2,3]	none	118/245 (48.2%)	101/199 (50.8%) 50.8%	OR 0.85 (0.58 to 1.25)	41 fewer per 1000 (from 133 fewer to 55 more) 41 fewer per 1000 (from 133 fewer to 55 more)	⊕⊕OO LOW	CRITICAL

[1] unclear reporting of allocation concealment
[2] the number of total events is low (less than 300)
[3] the CI crosses the line of no effect (includes 1) plus its lower limit is < than 0.75 and its upper limit is > than 1.25. In addition the total number of events is much lower than the rule of thumb of 300.

Question: Should more follow-up clinic visits versus fewer follow-up clinic visits be recommended for non metastatic colorectal cancer surveillance?

Quality assessment							Summary of findings					Importance
							Number of patients		Effect		Quality	
Number of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	More clinic visits	Fewer clinic visits	Relative (95% CI)	Absolute		
Overall survival at 5 years Jeffery 2007 (follow-up mean 5 years)												
2	randomised trials	serious[1]	serious[2]	no serious indirectness	serious[3,4]	none	116/394 (29.4%)	143/410 (34.9%) 34.9%	OR 0.78 (0.58 to 1.05)	54 fewer per 1000 (from 112 fewer to 11 more) 54 fewer per 1000 (from 112 fewer to 11 more)	⊕OOO VERY LOW	CRITICAL
Number of recurrences Jeffery 2007 (follow-up mean 5 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3,5]	none	123/394 (31.2%)	133/410 (32.4%) 32.4%	OR 0.93 (0.69 to 1.26)	16 fewer per 1000 (from 76 fewer to 53 more) 16 fewer per 1000 (from 75 fewer to 53 more)	⊕⊕OO LOW	CRITICAL

[1] unclear allocation concealment
[2] p value 0.1, I squared 62%
[3] the total number of events is low (less than 300)
[4] CI includes 1 and the lower limit is < 0.75
[5] CI includes 1, the lower limit is < 0.75 and the upper limit is >

Appendix 8

Publication 1

Kontoyannis A, Hargest R. How Guidelines influence modern surgical practice. In: Taylor I, Johnson CD (Editors). Recent Advances in Surgery. New Delhi: Jaypee Brothers Medical Publishers; 2011. p 20-33
(ISBN 978-93-5025-355-7)

Author Queries:
Please check Table 2.1, 2.2 and 2.3 is not cited.

CHAPTER TWO

How Guidelines Influence Modern Surgical Practise

Ms Angeliki Kontoyannis and Ms Rachel Hargest

INTRODUCTION

A medical guideline is a systematically developed statement to help clinicians and patients with decision-making regarding diagnosis, management, and treatment in specific areas of health care. The purpose is to standardise medical care to raise quality of care, reduce risk and achieve the best balance between cost and effectiveness.[1]

Guidelines have been in use for thousands of years during the entire history of medicine. In the late 5th century BC, Hippocrates wrote the Hippocratic oath. One of its main principles, "Do no harm" or non-maleficence, is today the cornerstone of medical ethics. Though originally intended to guide the practise of his pupils, the oath still holds relevance today and is taken by most new doctors across the world.[2]

In modern clinical practise, guidelines are gaining an ever-increasing presence. There are many sources of guidelines; governmental agencies, medical professional societies, health providers and patient groups. Opinion on the topic of guidelines differs amongst surgeons. There are those who believe that guidelines are a welcome development that can only bring improvement to clinical practise. On contrary, others believe that the guidelines are an unnecessary external authority that curbs the autonomy of individual surgeon. They question the validity of the development process and the quality of the body of evidence on which guidelines are often-based. There are also those who see the danger of a guideline becoming the only accepted option particularly in the context of a busy practise or a resource-strapped health provider. Others just feel overwhelmed by the sheer number of guidelines available, finding it, is often conflicting and confusing. For example, a search for guidelines on the management of heart failure resulted in over 1,000 citations.

The scope of this chapter is to address these points by providing an in depth look at guideline development and appraisal so that surgeons are encouraged to embrace this tool, which can be an invaluable piece of distilled medical information. However, as with everything in surgical practise one needs to learn to use it wisely, with thorough understanding

of the risks and benefits, always require judgment and surgical acumen before implementing it.

Guidelines are never mandatory by definition. However, there are situations where, by law, a doctor whose practise deviates from certain guidelines may be faced with legal proceedings.

STATUTORY GUIDELINES (SUBJECT TO LAW)

Guidance Produced by the General Medical Council (GMC)

The purpose of the GMC is to protect, promote and maintain the health and safety of the public by ensuring proper standards in the practise of medicine. The GMC was established under the Medical Act of 1858. Over time a range of new legislation has been introduced that defines their powers and responsibilities. The law gives the GMC four main functions under the Medical Act 1983:

- Keeping an up-to-date register of qualified doctors
- Fostering good medical practise
- Promoting high standards of medical education and training
- Dealing firmly and fairly with doctors whose fitness to practise is in doubt.

In 2010 the GMC merged with the Postgraduate Medical Education Training Board (PMETB) and has a much larger role in medical education and training.

To practise medicine in the UK, all doctors are required by law to be both registered with the GMC and hold a license to practice. Licensed doctors are required to demonstrate to the GMC that they are practising in accordance with the generic standards of practise laid down by them.[3]

Good Medical Practise (2006)[4] is the core guidance which the GMC produces for doctors regarding their fitness to practise. This guidance sets out the principles and values on which good practise and medical professionalism is founded. It covers the following domains:

- Duties of a doctor
- Principles of good clinical care
- Maintaining good clinical care
- Working with colleagues
- Probity
- Relationship with patients
- Teaching, training and appraisal
- Health of the doctor.

Quoting directly from the Good Medical Practise Document 2006 "It is the responsibility of every doctor registered with the GMC to be familiar with Good Medical Practise (2006) and to follow the guidance it contains. IT IS GUIDANCE, NOT A STATUTORY CODE, so every doctor must use their judgement to apply the principles to the various situations they are faced with. Every doctor must be prepared to explain and justify his or

her decisions and actions. Serious or persistent failure to follow this guidance will put your registration at risk."[4]

This is therefore a situation where by law a regulatory body has the power to remove a doctor's licence to practise, if the doctor fails to justify appropriately the reasons for deviating from the guidance.

ADVISORY GUIDELINES

Although the production of advisory guidelines does not directly relate to an act of parliament in the same way as the GMC guidelines do, failure of a doctor to comply with advisory guidelines can also lead to dispute with an employer or in case of harm or perceived harm due to negligence in proceedings. It is important to highlight that "a doctor is not guilty of negligence, if he has acted in accordance with a practise accepted as proper by a responsible body of medical men skilled in the relevant art" (Bolam test 1975).[5] Therefore, once again if doctors go against the accepted professional guidance than they should be able to justify their actions appropriately.

There are different bodies that produce advisory guidelines.

1. **Guidance Produced by Government** [e.g. Department of Health (DoH), Welsh Assembly, Scottish Parliament]

   These are usually service related guidelines (e.g. referral pathways, referral timelines). They are usually consensus statements that are the result of working groups, where invited specialist(s) have been asked to give their opinion(s) to policy makers. Adherence is often strongly recommended by employers (health providers) or government agencies, and financial incentives or penalties are often used to ensure compliance.

2. **Guidance Produced by Local Health Care Providers** (e.g. health care trusts or local networks)

These are both service and clinical guidelines. Typical examples are local prophylactic antibiotic prescription guidance, thrombosis prophylaxis guidance or blood transfusion guidance. These may be evidence-based or based on other published guidelines (from national guideline developers or professional bodies such as professional colleges/associations) with appropriate adaptation to the local community being served. They may also have consensus statements from local committees. These committees are usually made up of relevant specialist staff that volunteer their time to represent their department or specialty in the committee. Internationally there is a recent trend for local hospitals to employ professional guideline developers to oversee their guideline and protocol development (Australia, USA). Adherence is strongly recommended and certain employment contracts include a clause regarding possible legal action (under employment law) against the employee who fails to adhere to local protocols or agreed processes.

3. **Guidelines Produced by the National Institute of Health and Clinical Excellence (NICE)**

National Institute of Health and Clinical Excellence is an independent body commissioned by the Department of Health to produce guidelines for health care professionals treating patients in the National Health Service (NHS) in England and Wales. NICE guidelines are evidence-based recommendations designed to promote good health and prevent ill health. They are developed with transparent processes using the principles of evidence-based medicine. The guidelines address both clinical-effectiveness and cost-effectiveness issues.[6]

The stages of NICE guideline development are as follows:[7]

- The Department of Health asks NICE to produce a guideline on a particular topic (e.g. diagnosis and management of colorectal cancer)
- NICE commissions the appropriate National Collaborating Centre (NCC) to co-ordinate the development of the guideline for use in England and Wales. The seven National collaborating centres are independent centres responsible for the development of NICE guidelines. Each collaborating centre has a partnership of professional organisations, academic units and patient/care giver organisations. The technical team of professionals working at the NCC who support the guideline development consists of project managers, information specialists, health economists and reviewers. A management board oversees the guideline development work. The board comprises of representatives of relevant professional bodies. The management board monitors the operation of the NCC according to the contract with NICE, reviews the guideline development process and advises on changes to the guideline as negotiated with NICE at the end of the guideline development. The guideline team in the centre for clinical practise at NICE supports and advises the NCC during the development process. The main stages in the process are as follows:
- Agreeing the scope, what the guidelines will cover
- Establishing a Guideline Development Group (GDG) to manage the work. The GDG is composed of relevant health professionals, but also at least two patient/care giver representatives. GDGs usually consist of 12 to 15 people. NICE is not represented on the GDG
- Searching for, appraising and building research evidence
- Accessing and incorporating expert opinion when and if needed
- Developing the recommendations
- Consulting the views of stakeholder organisations on the provisional guidelines. Stakeholder organisations are organisations with an interest in a particular guideline. They register with NICE at the beginning of the process and contribute their views during the consultation period
- At the end of the process, the Guidance Executive at NICE signs off the guideline. The Guidance Executive confirms that the NCC

has developed the guideline in accordance with the terms of the remit from the Secretary of State for Health and the Scope, and by following NICE process and methods. The guideline is then published and distributed to the NHS in England and Wales.

There are different types of NICE guidelines, which are elucidated as follows:

- Clinical guidelines cover aspects of the management of a particular disease or condition. The evidence supporting different treatments is examined to assess whether they are effective for patients. The guidelines make recommendations on which treatments should be made available in the NHS in England and Wales, in order to ensure the best care is available to all patients. Clinical guidelines sit alongside, but do not replace the knowledge and skills of experienced health professionals and consider both the clinical effectiveness and also the cost-effectiveness of cancer treatments.
- Service guidelines make recommendations on how NHS services for patients should be organised in England and Wales. Both the anticipated benefits and the resource implications of implementing the recommendations are considered
- Technology appraisal guidance focuses on the clinical and cost effectiveness of one or more technologies, such as new drugs, surgical procedures and medical devices
- Interventional Procedures (IP) guidance covers the safety and efficacy of interventional procedures used for diagnosis or treatment
- Public health guidance deals with promoting good health and preventing ill health.

The Scottish Intercollegiate Guidelines Network (SIGN) develops evidence based clinical practise guidelines for the NHS in Scotland. SIGN guidelines are derived from a systematic review of the scientific literature. SIGN guidelines are produced by guideline development group members with support from the SIGN executive according to structured robust methodology.[8,9]

The Guidelines Audit and Implementation Network (GAIN) produce guidelines for the NHS in Northern Ireland. Its role is safety and quality improvement in Health and Social Care Services throughout Northern Ireland by commissioning of regional audit and guidelines as well as by the promotion of good practise through the dissemination of audit results and the publication, and facilitation of implementation of regional guidelines.[10]

4. **Guidelines Produced by International Guideline Developing Bodies**

Since the establishment of NICE in the UK, other countries are also establishing guideline-developing bodies. In the US there is an Agency for Health care Research and Quality.[11] In the Netherlands, two bodies, Community-Based Organization (CBO) and National Health care Group

(NHG) publish specialist and primary care guidelines respectively.[12] In Germany, the German Agency for Quality in Medicine (ÄZQ) co ordinates a national programme for disease management guidelines.[13] All these organisations are now, members of the Guidelines International Network (GIN), an international network of organisations and individuals involved in clinical practise guidelines.[12] GIN is the owner of the International Guideline Library: The largest web-based database of medical guidelines worldwide; and pursue a set of activities aiming at promoting best practise and reducing duplication in the guideline world. GIN also holds annual international conferences. The US and other countries also maintain medical guideline clearing houses. In the US, the National Guideline Clearing House maintains a catalogue of high-quality guidelines published by various organisations (mostly professional physician organisations).

5. **Guidelines Produced by Professional Medical Organisations and Societies**

Specialist working groups formed by members of the executive, who have volunteered to sit on the guideline panel, usually produce these guidelines. An alternative is that panel members are self-selected or nominated by their peers. The evidence provided varies with some societies producing very good quality evidence-based guidelines and others producing a higher number of consensus statements particularly when addressing topics where evidence is not available in the literature. A search of the UK-based societies relevant to surgical practise has found that the great majority of professional bodies and societies produce guidelines. However, only one society, i.e. British Society of Gastroenterology (BSG) gives reference to their detailed methodology and uses established, and recognised guideline development tools for their production. This is also the only organisation whose reports have a formal application process for guideline development panel members.

ASSESSMENT OF GUIDELINES

In order to assess the quality of a guideline the reader must be aware of the principles of evidence-based medicine and apply a systematic appraisal to the guideline. Detailed guides for assessing the validity of practise guidelines have been developed using rigorous methodology (see www.agreecollaborative.org).

The appraisal of a guideline should begin by identifying the scope and purpose of the guideline. This should be clear and the target users for whom the guideline has been written also need to be clearly identified.[14]

The main guideline document has two distinct components, which are as follows:

- The evidence summary
- The recommendations, a detailed instructions for applying the evidence to a specific population of patient.

TABLE 2.1

Illustrates that a lot of effort and resources are being placed in guideline development and it highlights the need for surgeons to be able to assess the quality of guidelines and judge their content before applying them to their practise irrespective of the source of their production

Society	*Produce guideline Yes (Y); No (N)*	*Methodology online or in guideline document Yes (Y); No (N)*	*Quality checklist used (AGREE tool) Yes (Y); No (N)*	*Nature of guideline work*
Association of Coloproctology of Great Britain and Ireland (ACPGBI)	Y	N	N	Regular guideline production on variety of surgical topics. Clinical and service guidelines
Association of Surgeons of Great Britain and Ireland (ASGBI)	Y	N	N	Regular guideline production on variety of surgical topics. Clinical and service guidelines
Association of Laparoscopic Surgeons of Great Britain and Ireland (ALSGBI)	Y	N	N	Occasional guidelines on laparoscopic topics
Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland (AUGIS)		N	N	Regular guideline production on variety of surgical topics. Clinical and service guidelines. Also refer to NICE and SIGN
British Association of Aesthetic Plastic Surgeons (BAAPS)	N	NA	NA	
British Association of Day Surgery (BADS)	Y	N	N	Occasional service guidelines. Some clinical
British Association of Otorhinolaryngologists (ENT-UK)	Y	N	N	Occasional guidelines on ENT topics
British Association of Plastic, Reconstructive and Aesthetic Surgeons (BAPRAS)	N	NA	NA	
British Association of Surgical Oncology (BASO) including Association of Breast Surgery(ABS~BASO)	Y	N	N	Several guidelines on the management of various aspects of breast cancer

Contd...

Contd...

Society	*Produce guideline Yes (Y); No (N)*	*Methodology online or in guideline document Yes (Y); No (N)*	*Quality checklist used (AGREE tool) Yes (Y); No (N)*	*Nature of guideline work*
British Association of Urological Surgeons	Y	N	N	Occasional guidelines on urological topics
British Hernia Society	N	NA	NA	
British Orthopaedic Association	Y	N	N	Regular guidelines on variety of orthopaedic topics
British Trauma Society	No access to website for non members	NA	NA	
Royal College of Physicians and Surgeons of Glasgow (RCPSG)	N	N	N	Provide links to SIGN
Royal College of Surgeons of Edinburgh (RCSE)	N	N	N	Provide links to SIGN
Royal College of Surgeons of England (RCS)	Y	N	N	RCS hosts the National Collaborating Centre for Acute Care (NCC-AC) producing NICE guidelines. It also has the Clinical Effectiveness Unit (CEU) producing documents that audit and disseminate best surgical practise
Royal College of Surgeons of Ireland (RCSI)	Y	N	N	First guideline under development on Quality Assurance in Radiology
Society for Cardiothoracic Surgery in Great Britain and Ireland.	No access to website for non members	NA	NA	
The Vascular Society of Great Britain and Ireland	Y	N	N	Occasional production of guide-lines on a variety of vascular topics

When assessing the quality of the evidence summary, it is important to check the following:

- All the relevant evidence has been identified through a thorough search method
- The review is recent enough or updated recently to be valid
- The inclusion and exclusion criteria for data extraction are clearly described
- The evidence has been graded for its validity
- The side-effects or risks of treatments are presented along with the benefits.

When assessing the recommendations it is important to consider whether:

- They have been graded for their validity
- The burden of illness, the guideline refers to, is significant enough in the population to warrant implementation
- The beliefs of the patients or the community are incompatible with the guideline
- The cost of implementation of the guideline is a good allocation of resources for the local community
- There are any insurmountable barriers (geographical, organisational, traditional, authoritarian, legal, or behavioural) to the implementation of the guideline.

Valid guidelines create their evidence components from systematic reviews of all the relevant worldwide literature. The reviews that provide the evidence components for guidelines are "necessity-driven", and synthesise the best available evidence. Therefore, some recommendations may be derived from evidence of high validity and others from evidence that is much more liable to error.[14]

Sources of evidence range from small laboratory studies or case reports to well-designed large clinical studies that have minimised bias to a great extent. Since poor quality evidence can lead to recommendations that are not in patients' best interests; it is essential to know whether a recommendation is strong (we can be confident about the recommendation) or weak (we cannot be confident).[15]

Grading schemes have been used for over 25 years. Since 1970s a growing number of organisations have employed various systems to grade the quality (level) of evidence and the strength of recommendations. Some grading systems are based on study design alone without explicit consideration of other important factors in determining quality of evidence. Some systems are excessively complex.[15]

A commonly used grading system categorizes evidence as follows:[16]

Unfortunately, different organisations use different systems to grade evidence and recommendations. The same evidence and recommendation could be graded as "II-2, B", "C+, 1", or "strong evidence, strongly recommended" depending on which system is used. This is confusing and impedes effective communication.[15]

TABLE 2.2

Scheme for grading evidence and recommendations

	Grading of evidence
Ia	Evidence obtained from meta-analysis of randomised controlled trials
Ib	Evidence obtained from at least one randomised controlled trial
IIa	Evidence obtained from at least one well-designed controlled study without randomisation designed quasi-experimental study
III	Evidence obtained from a well-designed non-experimental descriptive study, such as comparative studies, correlation studies and case studies
IV	Evidence obtained from expert committee reports or opinions or clinical experiences of respected authorities
	Grading of recommendations
A	Evidence categories Ia and Ib
B	Evidence categories IIa, IIb and III
C	Evidence category IV

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group began in the year 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in health care. The working group has developed a common, sensible and transparent approach to grading quality of evidence and strength of recommendations. Many International Organisations have provided input into the development of the approach and have started using it. The GRADE system is used widely nowadays. The World Health Organisation, the American College of Physicians, the American Thoracic Society, up-to-date (an electronic resource widely used in North America; http://www.uptodate.com), NICE and the Cochrane Collaboration are among the organisations that have adopted GRADE.[17]

GRADE SYSTEM CLASSIFICATION FOR QUALITY OF EVIDENCE[14]

To achieve transparency and simplicity, the GRADE system classifies the quality of evidence in one of four levels:

Quality of Evidence and Definitions

- *High quality*: Further research is very unlikely to change our confidence in the estimate of effect
- *Moderate quality*: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
- *Low quality*: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
- *Very low quality*: Any estimate of effect is very uncertain.

Some of the organisations using the GRADE system have chosen to combine the low and very low categories. Evidence-based on randomised controlled trials begins as high quality evidence, but our confidence in the evidence may decrease because of several reasons, including:

- Study limitations
- Inconsistency of results
- Indirectness of evidence
- Imprecision
- Reporting bias.

Although observational studies (for example, cohort and case-control studies) start with a "low quality" rating, grading upwards may be warranted, if the magnitude of the treatment effect is very large, if there is evidence of a dose-response relation or if all plausible biases would decrease the magnitude of an apparent treatment effect.

GRADE SYSTEM CLASSIFICATION FOR STRENGTH OF RECOMMENDATIONS[14]

For recommendations, the GRADE system offers two grades:

- Strong
- Weak.

When the desirable effects of an intervention, clearly outweigh the undesirable effects or do not, guideline panels offer strong recommendations. On the other hand, when the trade-offs are less certain (either because of low quality evidence or because evidence suggests that desirable and undesirable effects are closely balanced), weak recommendations become mandatory.

Advantages of GRADE over other systems are as follows:[14]

- Developed by a wide representative group of international guideline developers
- Clear separation between quality of evidence and strength of recommendations
- Explicit evaluation of the importance of outcomes of alternative management strategies
- Explicit comprehensive criteria for downgrading and upgrading quality of evidence ratings
- Transparent process of moving from evidence to recommendations
- Explicit acknowledgment of values and preferences
- Clear pragmatic interpretation of strong versus weak recommendations for clinicians, patients, and policy makers
- Useful for systematic reviews and health technology assessments, as well as guidelines.

TABLE 2.3

Summary for quality assessment of a guideline

- The methodology of the guideline development must be robust and clearly presented
- The search dates, engines, databases must all be clearly presented
- The level of evidence must be clear and presented next to each recommendation made
- GRADE is a new system of grading evidence and guideline developers are advised to use this system
- Any cost analysis should be explicit and the economic evidence on which the model has been based must also be graded and presented clearly.

CONCLUSION

Guidelines are becoming an ever-increasing reality in modern surgical practise. Their purpose is to improve quality of health care provision. With a variety of governmental and medical organisations producing guidelines, it is imperative that the modern surgeon is educated in the principles of evidence-based medicine. This way each guideline is assessed on its own methodology and development process in the setting, is going to be implemented. It is possible that the same body of evidence will produce different guidelines when different guideline developers based on different population or setting. It is always down to the individual surgeon to decide on the individual patient's management. Guidelines aim to assist in this process but never replace the medical acumen and experience of the specialist or the wishes of the patient.

REFERENCES

1. Institute of Medicine. Clinical Practice Guidelines: directions for a new program. Washington, DC: National Academy Press 1990.
2. Edelstein L. The Hippocratic oath: text, translation and interpretation. Baltimore: Johns Hopkins University Press 1967; p. 56.
3. General Medical Council (2011) [online]. GMC website. Available from http://www.gmc-uk.org/licencing/index.asp [Accessed 2011].
4. General Medical Council (2011) Good Medical Practice [online]. GMC website. Available from http://www.gmc-uk.org/good_medical_practice/index.asp [Accessed 2011].
5. Jones, M. Medical Negligence. 3rd edition. London: Sweet and Maxwell 2003.
6. National Institute for Health and Clinical Excellence (2011) About NICE [online]. NICE website. Available from http://www.nice.org.uk/aboutnice [Accessed 2011].
7. National Collaborating Centre for Cancer (2011) Developing Cancer Guidelines [online]. NCCC website. Available from http://wales.nhs.uk/sites3/page.cfm?orgid=432 and pid=12489 [Accessed 2011].

8. Scottish Intercollegiate Guidelines Network (2011) About SIGN [online]. SIGN website. Available from http://www.sign.ac.uk/about/index.html [Accessed 2011].
9. Scottish Intercollegiate Guidelines Network (2011) Methodology [online]. SIGN website. Available from http://www.sign.ac.uk/methodology/index.html [Accessed 2011].
10. Guidelines and Audit Implementation Network (2011) Welcome to gain-ni.org. [online]. GAIN website. Available from http://www.gain-ni.org [Accessed 2011].
11. National Guideline Clearinghouse (2011) Public resource for evidence-based clinical practice guidelines [online]. Agency for Healthcare Research and Quality's (AHRQ) website. Available from http://guidelines.gov [Accessed 2011].
12. Guidelines International Network (2011) [online]. Guidelines International Network website. Available from http://g-i-n.net [Accessed 2011].
13. Agency for Quality in Medicine (2011) [online]. Agency for Quality in Medicine (AQuMed/ÄZQ)website. Available from http://aqumed.de [Accessed 2011].
14. Strauss ES, Richardson WS, Glasziou P, Haynes RB. Evidence-Based Medicine: How to practice and teach EBM. 2nd edition. Edinburgh: Elsevier-Churchill Livingstone p. 165-66.
15. Gordon H Guyatt, Andrew D Oxman, Gunn E Vist, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.
16. Cairns SR, Scholefield JH, Steele RJ, et al. Guidelines for colorectal cancer screening and surveillance in moderate and high-risk groups. Gut 2010;59(5):666-90.
17. GRADE working group. (2011) [online]. GRADE working group website. Available from *www.gradeworkinggroup.org* [Accessed 2011].

Appendix 9

Publication 2

Kontoyannis A, Hargest R. The importance of understanding guideline methodology and the principles of evidence based medicine. BMJ online rapid response. 2011. http://www.bmj.com/rapid-response/2011/11/03/importance-understanding-guideline-development-methodology-and-principles-

Analysis

Breaking the rules: understanding non-compliance with policies and guidelines

BMJ 2011; 343 doi: http://dx.doi.org/10.1136/bmj.d5283 (Published 13 September 2011) Cite this as: BMJ 2011;343:d5283

- Article
- Related content
- Article metrics
- Rapid responses
- Response

The importance of understanding guideline development methodology and the principles of evidence-based medicine.

We enjoyed reading the article by Carthey et al. regarding the use of guidelines in clinical practice. The authors have presented an important issue and their frustration is undoubtedly shared by many clinicians.

To accept the recommendations of any guideline and potentially change practice individual clinicians need the skills to appraise the validity of the guideline.

The current confusion of many healthcare professionals with regard to guidelines that has been reported by the authors, is accentuated by a lack of appraisal skills.

We may eagerly read the recommendations, but this is incomplete if we don't read or understand the methodology in order to assess the quality and validity of the guideline.

Our research shows that there is very poor reporting of methodology by the majority of UK learned medical societies that produce guidelines. The methodology is often absent from the guideline document and the society website.[1] Even when the methodology is reported, has been assessed and is sound, clinicians need to be able to assess the strengths and weaknesses of individual recommendations. Contradictory statements between different guidelines are likely to arise when the evidence is of low quality and the recommendation has been developed by consensus methods. If the evidence is of high quality there may be subtle differences in the development framework because focusing on subtly different end points can produce a very different search result and therefore recommendation.

Differences between international guidelines may reflect the

different healthcare settings and culture of both the patient population and the guideline development group. These should be accounted for when assessing the applicability of the guideline to our own patients.

We agree with the authors that guidelines should be a means to improve patient care. To streamline the use of what we consider a valuable new tool and navigate through this currently ever-increasing ocean of information we would also like to see:

1.better reporting of methodology by learned medical societies on production of their guidelines.

2.education in guideline development, appraisal and the principles of evidence-based medicine as a core curriculum topic in medical schools and postgraduate training programmes.

Reference:

1.Kontoyannis A, Hargest R. How guidelines influence modern surgical practice. In Recent Advances in Surgery. Ed:Taylor I, Johnson CD. Jaypee London 2011:20-33.

Competing interests: No competing interests

06 October 2011

Angeliki Kontoyannis
ST4 Registrar London Deanery, previously research fellow National Collaborating Centre for Cancer (N
Rachel Hargest, Consultant Surgeon Cardiff University / Royal College of Surgeons of England representative on NCC-C management board
The NCC-C develops NICE cancer guidelines for England and Wales

Appendix 10

Publication 3

Kontoyannis A, O'Connell, S, Berendse S, Poston G, Hargest R. Colorectal liver metastases - which diagnostic imaging modality is best. British Journal of Surgery. 2013;100 (S4):2–49

Presented at the Joint Conference of the Society for Research and Academic Surgery and the Royal Society of Medicine, RSM London, Jan 2013.

O67

Oncological feasibility of laparoscopic distal pancreatectomy for adenocarcinoma of distal pancreas: A single-institution comparative study

S. Rehman, R. Lochan, J. J. French, B. C. Jaques, D. M. Manas, S. A. White

Department of Hepatobiliary Surgery, Freeman Hospital

Introduction: Laparoscopic distal pancreatectomy (LDP) is standard procedure for resection of distal pancreatic pathology. However, concerns have been raised regarding its oncologic adequacy in treating pancreatic malignancies as compared to open distal pancreatectomy (ODP). The aim of this study was to explore this issue.
Methods: We retrospectively analysed a prospectively maintained Unit database for patients who underwent distal pancreatic resection between Jan 2005 and March 2012. Data was analysed on SPSS® v19 utilising standard tests. A p-value of <0.05 was considered significant.
Results: Of 130 patients who underwent DP, n = 57 had histologically confirmed adenocarcinoma (LDP n = 23, ODP n = 34). M: F were 12:11 vs 14:20, mean age was 64 vs 62 years (p = 0.68) and mean (95% CI) pathological size of the lesion size was 27 (18.2 - 36.9) mm and 36 (28–44.2) mm p = 0.16 in LDP and ODP respectively. Mean (CI 95%) intraoperative blood loss in LDP = 583 (146.7-1019.4) mls was significantly lower (p = 0.05) compared to ODP = 1396 (765–2027.6) mls, but operative time was longer (339 (299.3-379) mins vs. 235 (202–268) mins (p = 0.00). LDP resulted in non-significant shorter epidural requirements (3.61hrs (1.8-9hrs) vs. 18.54 hrs (6-31hrs), (p = 0.06). Postoperative pancreatic fistula formation was similar between [LDP n = 4 (17 %), ODP n = 7 (20 (p = 0.35)]. Importantly the R0 rates were similar in both groups [(R0/R1 resection LDP = 21/2, ODP = 29/5 (p = 0.61)]. No difference in peri-operative mortality and 1-year survival was seen.
Conclusion: LDP is results in oncologically comparable results as compared to ODP for adenocarcinoma arising in the distal pancreas.
Take-home message: laparoscopic resection for distal pancreatic adenocarcinoma provides oncological results comparable to the open procedure.

O68

Anatomical variation in the inferior pancreaticoduodenal arteries and its relevance to pancreas transplantation

R. Al-biatty

University of Leeds

Introduction: Knowledge of the variation in anatomy of the IPDaa is important in pancreas transplantation as it can potentially complicate procurement and revascularization of the pancreas. The aims of this study were to investigate the variation in the patterns of origin of the inferior pancreaticoduodenal arteries (IPDaa.), and the variation in the level at which the IPDaa arise from the superior mesenteric artery (SMA).
Methods: The abdomens of 20 cadavers (10 males; 10 females) were dissected to demonstrate the range of variation in patterns of origin of the IPDaa. The incidence and diversity of the variations were recorded. The distance from the aorta to the level of origin of the IPDaa as they arise from the SMA was measured.
Results: 9 different patterns of origin were found for the anterior IPD artery (IPDa), and 10 different patterns of origin were found for the posterior IPDa. The results of measuring the distance from the aorta to the origin of the IPDaa ranged from 24-57mm.
Conclusions: The pattern of origin of the IPDaa is extremely variable with clinical relevance in pancreas transplantation. It has been reported that in some cases, the head of the pancreas does not reperfuse during the reperfusion stage of transplantation. This arises when ligation of the SMA occurs above the origin of the IPDaa. The IPDaa are the only arterial supply to the head of the donor pancreas. Thus, knowledge of the variation in anatomy of the IPDaa will obviate extensive dissection to identify vessels, and avert vascular damage.
Take-home message: This study, with the use of cadaveric dissections, has shown that variation does exist in the patterns of origin of the inferior pancreaticoduodenal arteries, and that variation also exists in the level at which the inferior pancreaticoduodenal arteries arise from the SMA. This knowledge is important because the single most important detail during a multi-organ procurement is retention of the inferior pancreaticoduodenal arteries with the pancreas as it is essentially the only bloody supply to the head of a transplanted pancreas.

O69

Colorectal liver metastases – which diagnostic imaging modality is best?

A. Kontoyannis[1,2,3], S. O'Connell[1], S. Berendse[1], G. Poston[1,4] R. Hargest[1,2,5]

[1]National Collaborating Centre for Cancer, [2]Cardiff University, [3]University Hospital Lewisham, [4]University Hospital Aintree, [5]University Hospital Wales

Introduction: Up to 20% of patients scheduled for resection of colorectal liver metastases have unresectable disease at laparotomy. Following staging CT, MRI and PET-CT scans further assess liver lesion resectability. Often patients have both tests. PET-CT is considered more accurate but is expensive and less widely available. MRI is more available and cheaper. Our aim was to perform a systematic review and meta-analysis of the diagnostic capabilities of these imaging modalities.
Methods: Databases including MEDLINE and EMBASE were searched. Study quality appraisal was carried out using the QUADAS tool for diagnostic studies. Meta-analysis was performed using RevMan5 software.
Results: Of 287 papers identified 22 were selected for review. 77% of studies did not report blinding. All studies had multiple reference tests. Per-patient analysis reports summary sensitivity and accuracy for PETCT as 94% for both and is higher than MRI (80%,91%) and CT(87% for both). Per-lesion analysis shows MRI reporting higher sensitivity compared to CT and PETCT with MRI pooled sensitivity of 88% and accuracy of 87%, CT a sensitivity of 74% and accuracy of 78% and PETCT a sensitivity of 79% and accuracy of 97%.
Conclusions: The quality of the studies is low with high risk of bias mainly due to varied study design and poor reporting of design parameters. On the basis of current evidence it is unclear whether MRI or PET-CT should be used after CT scan to confirm whether the patient with liver metastases is suitable for surgery.
Take-home message: On the basis of current evidence it is unclear whether MRI or PET-CT should be used after CT scan to confirm whether the patient with liver metastases is suitable for surgery.

O70

Identification of molecular pathways implicated in the pathogenesis of oxaliplatin induced sinusoidal obstruction syndrome

S. M. Robinson[1,2], J. Mann[1], D. M. Manas[2], D. A. Mann[1], S. A. White[1,2]

[1]Institute of Cellular Medicine, Newcastle University, [2]Department of HPB Surgery, Freeman Hospital, Newcastle upon Tyne

Introduction: Oxaliplatin based chemotherapy is widely utilised pre-operatively in patients with colorectal liver metastases. Its use has been associated with an injury to the liver in the form of sinusoidal obstruction syndrome (SOS) the presence of which can have a negative impact on surgical outcomes. At present the pathogenesis of this condition is poorly understood.
Methods: C57Bl/6 mice (n = 10 per group) were treated with 5-FU/Oxaliplatin/Folinic Acid chemotherapy (FOLFOX) weekly for 5 weeks or their respective vehicle controls. Animals were culled one week following the final treatment and liver tissue harvested for histological and biochemical analysis. mRNA was extracted from snap frozen liver and subject to