Diagnostic Accuracy of MRI for Assessment of T-category, Lymph Node Metastases, and Circumferential Resection Margin Involvement in Patients with Rectal Cancer: A Systematic Review and Meta-analysis

by

Eisar Al-Sukhni

A thesis submitted in conformity with the requirements
for the degree of Master of Science (Clinical Epidemiology)

Institute of Health Policy, Management and Evaluation
University of Toronto

DIAGNOSTIC ACCURACY OF MRI FOR ASSESSMENT OF T-CATEGORY, LYMPH NODE METASTASES, AND CIRCUMFERENTIAL RESECTION MARGIN INVOLVEMENT IN PATIENTS WITH RECTAL CANCER: A SYSTEMATIC REVIEW AND META-ANALYSIS

Eisar Al-Sukhni

Master of Science 2012

Institute of Health Policy, Management and Evaluation

University of Toronto

Abstract

BACKGROUND: MRI is increasingly being used for rectal cancer staging. The purpose of this study was to summarize published evidence to determine the accuracy of MRI for T-category, lymph node (LN) metastases, and circumferential resection margin (CRM) involvement in rectal cancer.

METHODS: Sensitivity, specificity, and diagnostic odds ratios (DOR) were estimated using hierarchical summary receiver operating characteristics modeling and bivariate random effects modeling.

RESULTS: MRI was more specific for CRM (94%, 95%CI 88-97) than for T-category (75%, 95%CI 68-80) and LN's (71%, 95%CI 59-81) but was more sensitive for T-category (87%, 95%CI 81-92) than for CRM (77%, 95%CI 57-90) and LN's (77%, 95%CI 69-84). DOR was higher for CRM (56.1, 95%CI 15.3-205.8) than for LN's (8.3, 95%CI 4.6-14.7) and T-category (20.4, 95%CI 11.1-37.3).

CONCLUSIONS: MRI has good accuracy for both CRM and T-category and should be considered for preoperative rectal cancer staging. In contrast, LN assessment is poor on MRI.

Acknowledgments

I extend my sincerest gratitude to my parents for their support and encouragement through all my educational endeavours. My thanks also go to Dr Joseph Beyene who contributed his expertise and provided invaluable assistance with statistical analysis. Last but not least, I would like to thank the members of my thesis committee for their supervision and guidance throughout the course of my graduate work.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1
Introduction

1.1 Rectal cancer and the problem of local recurrence

Rectal cancer is the seventh most commonly diagnosed cancer in Canada, with 6,721 new cases diagnosed in 2007 (1). Treatment is primarily surgical supplemented by chemotherapy and/or radiation for improved disease control. Using the American Joint Committee on Cancer (AJCC) TNM pathology classification, rectal cancer is described according to the depth of tumour invasion through the layers of the rectal wall (T), number of regional lymph nodes involved (N), and the presence or absence of distant metastases (M) (TABLE A.1a). Combinations of T, N, and M categories are grouped into stages (I-IV) (TABLE A.1b) which are used to predict prognosis and guide decision-making with respect to the need for further treatment.

Despite undergoing curative resection for rectal cancer, 5-11% of patients develop local recurrence of their disease within five years (2-4). Local recurrence is associated with a poor prognosis with an overall five year survival of only 9-14% (5,6). Morbidity can be significant as patients may suffer from unremitting rectal pain that is poorly controlled by narcotics and which reduces quality of life (7,8).

Several histopathological factors have been associated with increased risk of local recurrence, including advanced T-category (9-12), mesorectal lymph node involvement (11-14), circumferential resection margin (CRM) involvement (15-17), extramural vascular invasion (13), and poor tumour differentiation (16). The most important of these factors is CRM, defined as the shortest distance from the main tumour, any tumour deposit, or malignant lymph node to the mesorectal fascia which envelopes the rectum radially. Traditionally, routine pathologic assessment of rectal cancer specimens for CRM involvement consisted of sampling a variable number of blocks from the luminal surface. In their landmark paper published in 1986, Quirke and colleagues prospectively examined serial thin-section transverse slices of 52 resected rectal cancer specimens and demonstrated that CRM involvement could be identified in 14 specimens (27%) using serial slicing technique but only in 6 specimens (12%) using routine sampling (18). On follow-up, positive CRM was associated with significantly higher local recurrence compared

with a negative CRM (85 versus 3%, p<0.001). The "Quirke technique" became the standard practice to examine specimens for CRM involvement, and population-based studies have since shown that CRM is the single strongest predictor of local recurrence, more significant than T-category or lymph node metastases (15,16). A pathological CRM of 1 mm or less has been shown to predict a significantly higher rate of local recurrence and defines a positive margin (15,18,19,20).

1.2 The role of surgical technique

While CRM has been recognized as a critical determinant of local recurrence, achieving a clear margin remains somewhat of a challenge in practice. The single most critical modifiable variable in achieving a negative CRM is surgery; however, the fixity of the rectum deep within the confines of the bony pelvis and its proximity to other organs in a limited space make surgical removal of the rectum a technically difficult operation. Traditionally, rectal cancer surgery involved mobilization of the rectum and its surrounding mesorectum using blind, blunt dissection along undefined planes. This technique had high potential to leave behind residual mesorectum with viable tumour in the pelvis (21), resulting in a positive CRM. Consequently, local recurrence rates with conventional surgery were high, up to 30-40% (22,23).

In the early 1980's, the surgical technique of total mesorectal excision (TME) was introduced, whereby the entire rectum and mesorectum are removed using sharp dissection under direct vision along the mesorectal fascia on all sides (24). In contrast to conventional technique, this operation conforms to principles of oncologic surgery which dictate removal of the tumour and any involved surrounding structures in one complete package. Correctly performed, TME proceeds along defined avascular planes and avoids perforation of the mesorectum which can lead to tumour spillage and contamination and increase the risk of disease dissemination. The application of this technique significantly improved the possibility of achieving a negative CRM and thus reduced the incidence of local recurrence (3,25,26). In an analysis of observational studies from a number of countries, prospective outcomes of 691 patients with Stage II or III rectal cancer who underwent TME surgery at three specialized oncology centres (US, UK, Japan) in 1978-1994 were compared with 720 similar stage patients who underwent conventional surgery at multiple Norwegian and Dutch community hospitals in 1986-1993 (27). Actuarial

five-year local recurrence was 39-42% after conventional surgery and 4-9% after TME. A Cox proportional hazards model was used to test for differences in the risk of local recurrence between TME and conventional surgery groups adjusting for age, gender, tumour stage, tumour height, type of operation (low anterior or abdominoperineal resection), and adjuvant radiation and chemotherapy. Compared to the Japanese TME group, the relative risk of local recurrence was not significantly different in the other TME groups (RR 1.04, 95%CI 0.46-2.38 and 0.58, 95%CI 0.20-1.72) but was significantly higher in the two conventional surgery groups (RR 5.12, 95%CI 2.73-9.62 and 6.46, 95%CI 3.46-12.08).

While TME has become the standard of care in rectal cancer surgery, it is not always attempted successfully. The quality of TME surgery can be evaluated using the macroscopic appearance of the resected mesorectum specimen as a surrogate marker of completeness of resection. In the grading system established by Quirke and colleagues, a TME specimen is deemed "complete", "nearly complete", or "incomplete" based on the regularity and depth of any defects in the mesorectal surface and the degree of dissection into mesorectum ("coning") (18,28). Technical variables such as a narrow pelvis or obesity can exacerbate the difficulty of achieving adequate rectal dissection and increase the likelihood of incomplete TME (21). In a prospective study of 130 patients who underwent curative resection with attempted TME, positive CRM was significantly higher in patients with incomplete TME (29.4%) compared with complete (1.6%) and nearly complete (5.8%) TME (p=0.0001). Local recurrence rose in parallel, from 1.6% and 5.7% respectively after complete and nearly complete TME to 41% after incomplete TME (p=0.0001) (29). The relationship between TME quality, positive CRM, and local recurrence was confirmed by data from the international multicentre CR07 trial which compared the impact of short course preoperative radiotherapy versus selective postoperative chemoradiotherapy on outcomes of patients with resectable Stage I-III rectal cancer (30). In this study, the plane of surgery achieved with TME was similarly distributed in both arms and was strongly associated with positive CRM (9% for complete, 12% for nearly complete, 19% for incomplete TME, p=0.001) as well as local recurrence at 3 year follow-up (4% for complete, 7% for nearly complete, 13% for incomplete TME, p=0.0039). These results underline the importance of complete surgical resection as the key to reducing local recurrence.

1.3 The role of adjuvant therapy

Even when TME is performed adequately, the risk of local recurrence can be further reduced with the addition of radiation. Several major randomized controlled trials (RCTs) have investigated the impact of radiation therapy (RT) on local recurrence and survival following curative resection of rectal cancer (TABLE A.2). The only trial to demonstrate a survival benefit was the Swedish Rectal Cancer Trial (31) which compared patients who received surgery alone with those who received short-course RT prior to surgery. At five year follow-up, patients who received preoperative RT had significantly fewer local recurrences (11% versus 27%, p<0.001) and significantly higher overall survival (58% versus 48%, p=0.004) compared with the surgery alone group. A similar effect on local recurrence was demonstrated in the Dutch TME trial (2,32) which randomized patients with Stage I-III rectal cancer to receive TME surgery with or without preoperative short-course RT. At a median follow-up of six years, local recurrence among patients who underwent a macroscopically complete resection was significantly lower in patients who received preoperative RT than those who did not (5.6% versus 10.9%, p<0.001); however, overall survival was the same in both groups (64.2% and 63.5%, respectively). On multivariable analysis, the reduction in local recurrence risk was found to apply to Stage II and III (i.e. T3/T4 or lymph node positive) patients only. This study definitively demonstrated the benefit of preoperative RT in reducing local recurrence beyond that achieved by complete resection in selected patients.

Three subsequent RCTs studied the impact of adding chemotherapy to radiation in patients with Stage II and III rectal cancer who received TME surgery (33-35). One trial (33) compared preoperative short-course RT with preoperative chemoradiotherapy (CRT) and found no significant difference in local recurrence (9.0% versus 14.2%, p=0.17) or survival between the two groups, while the other two trials compared preoperative long-course RT with preoperative CRT and showed a significant reduction in local recurrence with the use of CRT ((8.1% versus 16.5%, p<0.05) (34) and 8.7% versus 17.1%, p=0.002 (35)) but no difference in overall survival.

The optimal time to administer CRT was investigated in a German trial (36) which compared local recurrence and overall survival in Stage II and III rectal cancer patients randomized to receive CRT before or after TME surgery. The rate of local recurrence at five year follow-up

was significantly lower among patients who received preoperative CRT compared with postoperative CRT (6% versus 13%, p=0.006), but overall survival was similar in the two groups. This study and others (36,37) also demonstrated that administration of radiation preoperatively rather than postoperatively results in higher efficiency per dose of radiation (37), lower toxicity to surrounding healthy tissue (37), and fewer acute and long-term side effects (36). Preoperative CRT also produces downstaging of tumours and may permit the use of less radical procedures, particularly in low rectal tumours (36,38). Current North American guidelines recommend preoperative CRT for all patients with Stage II and III rectal cancer (39,40).

Unfortunately, the broad use of preoperative RT or CRT poses certain challenges. First, despite their important benefits in reducing local recurrence, RT and CRT are associated with both acute and long-term adverse effects. In the Dutch trial, short-course RT was associated with increased operative blood loss (1100 versus 1000 mL, p<0.001) and perineal complications (48% versus 41%, p=0.008) (41), as well as slower recovery from defecation problems and increased sexual dysfunction (42) compared with patients who received no RT. On long-term follow-up, patients who received RT reported increased fecal incontinence (62% versus 38%, p<0.001), pad wearing as a result of incontinence (56% v 33%, p <0 .001), anal blood loss (11% versus 3%, p<0.004), anal mucus loss (27% versus 15%, p<0.005), and decreased satisfaction with bowel function (43). Increased fecal incontinence was similarly reported in the Swedish trial as a late adverse effect of preoperative RT (57% versus 26%, p=0·013), along with an increase in urinary incontinence (45% versus 27%, p=0.023) and cardiovascular side effects (35% versus 19%, p=0·032) (44). The addition of chemotherapy to preoperative RT resulted in increased (unspecified) grade 3 or 4 toxicity compared with short-course RT alone in the Polish trial, but no differences were reported in quality of life or anorectal or sexual function at a median of one year following surgery (45).

A second problem with administering preoperative CRT to all Stage II and III rectal cancer patients is that although the majority of rectal cancers present as T3 tumours (i.e., extending past the outer muscular layer of the bowel wall), as a group these tumours are the most variable with respect to survival and recurrence of disease. Research has shown that with surgery alone, early T3 tumours (i.e., where depth of invasion is ≤5mm) have a prognosis similar to T2 tumours, while those extending >5 mm deep have a poorer prognosis that is similar to T4 tumours (46).

These observational data were based on specimens resected prior to the institution of TME surgery as the gold standard and may therefore be related to inadequate resection of the CRM in the advanced tumours. Regardless, they have led some researchers to question the current practice of radiating all T3 tumours, arguing that patients with early T3 tumours may be exposed to the toxic effects of radiation without additional benefit (47-49).

1.4 The need for accurate preoperative staging

As the evidence reviewed above makes clear, modern management of rectal cancer is complex and varies depending on disease stage and depth of invasion. In the era of preoperative CRT, optimal patient outcomes dictate that much of treatment decision making be made preoperatively. While understaging rectal cancer may lead to omission of preoperative RT and CRT and an increased risk of local recurrence, overstaging may lead to unnecessary treatment with preoperative RT and CRT with significant functional consequences. Therefore, accurate preoperative staging is crucial in order to identify patients who would most benefit from CRT and minimize unnecessary morbidity.

1.5 Approaches to preoperative staging of rectal cancer

To date, the majority of preoperative therapy trials have used some combination of digital rectal examination (DRE), computed tomography (CT) and endorectal ultrasound (ERUS) to stage rectal cancers and assign treatment groups (TABLE A.2) (36,37,50).

DRE involves digital examination of the rectal tumour by the surgeon to assess its mobility and height above the anal verge. These factors are clinical indicators of the favourability of prognosis and are used to help determine the need for preoperative therapy. However, DRE is subjective and cannot provide precise staging information even in the hands of an experienced colorectal surgeon. In a study which evaluated the effectiveness of DRE in selecting patients for preoperative RT based on signs of locally advanced disease (fixed versus mobile tumour) compared with histologic assessment of the same (T4 or CRM$\leq$1mm versus T1, T2, or T3<1mm, lymph node negative), DRE had very poor agreement with histology (weighted kappa=0.16) and allocated patients to appropriate treatment groups in only 40% of cases (51). This mode of staging is also limited by physical factors; in the aforementioned study, DRE could not be

performed in 24/98 (25%) of eligible patients because tumours were too high in the rectum to be reached or the examination was too painful.

CT provides cross-sectional imaging of the entire abdomen and pelvis and allows visualization of the tumour in relation to surrounding structures. Its role in preoperative staging of rectal cancer is primarily in the detection of regional and distant metastatic spread. While it has been extensively investigated for local rectal cancer staging, CT lacks the resolution to depict depth of bowel wall penetration and consistently compares poorly with ERUS and MRI in T-category assessment (52,53) (see Section 2.1 for more detail); as a result, it is not generally recommended for this purpose (54,55).

ERUS is a dynamic imaging procedure performed by insertion of an ultrasound probe into the rectum. It displays the individual layers of the bowel wall as well as the mesorectum to demonstrate tumour depth of invasion and spread to mesorectal lymph nodes. ERUS has the advantage of positioning the image probe immediately adjacent to the tumour with minimal interference from interposing tissue. It provides excellent resolution of the bowel wall, particularly for early rectal tumours (T1 or T2) (see Section 2.1 for more detail) and is useful to identify patients who are eligible for local (transanal) tumour excision. ERUS is the current gold standard for staging rectal cancer, particularly in American guidelines (54,55). However, it is operator dependent and, like DRE, may not be technically feasible for proximal or stenotic rectal tumours (56).

Over the past decade, magnetic resonance imaging (MRI) has been extensively studied for staging rectal cancer and its quality has improved (57,58). While ERUS is more specific than MRI for distinguishing early rectal cancers (53), MRI has been shown to be superior for demonstrating the extent of deeper tumours and additionally better delineates anatomy than other imaging modalities (58). More importantly, MRI can visualize the mesorectal fascia to predict the potential CRM (59) which cannot be determined by either ERUS or DRE. In retrospective studies, CRM identified on MRI has been shown to be a good predictor of local recurrence (60,61), although it has not been used in any RCT as a determinant of preoperative treatment decision making. While T-category and lymph node involvement are primarily used to allocate

preoperative CRT in North America, predicted CRM status is a key element in several European guidelines for rectal cancer staging and preoperative treatment decision-making (55).

Chapter 2
Rationale

2.1 The need for a meta-analysis

Numerous studies have investigated the diagnostic accuracy of MRI for predicting T-category, lymph node involvement, and CRM, resulting in a wide range of estimates. In this situation, meta-analysis of individual diagnostic study data using formal statistical methods is helpful as it can determine how well a test classifies individuals with and without disease and how these estimates compare across tests (i.e., staging modalities). Meta-analysis has the advantage of taking into account all available evidence (including conflicting results) to answer one or more specific questions. Because a meta-analysis incorporates more data than any individual study, it may have more power to detect true differences in performance between tests than single studies. Furthermore, it can quantify the variability of test accuracy and the degree to which different factors can influence it (62).

In the past decade, four meta-analyses have investigated the accuracy of MRI for staging rectal cancer in comparison with histopathology as the gold standard (TABLE A.3). In a systematic Medline-based review, Kwok et al (52) compared ERUS, CT, and MRI for assessing T-category (penetration through muscularis propria versus no penetration) and lymph nodes (nodes involved versus uninvolved). The analysis involved 78 papers published 1981-1998 and included different types of MRI coils (body, endorectal, phased array). Patients who received preoperative RT were included in the analysis. Simple descriptive pooled analysis (see Chapter 3.3) was used to produce estimates of sensitivity and specificity of each imaging modality for T-category and lymph nodes. Sensitivity and specificity of ERUS, CT, and MRI were respectively: (93%,78%), (78%,63%), and (86%,77%) for T-category and (71%,76%), (52%,78%), (65%,80%) for lymph nodes. Since preoperative long-course RT can cause tumour regression (thus downstaging tumours) and alter accuracy estimates, separate subgroup analyses were performed for patients who did not receive preoperative RT. In this smaller group of patients, sensitivity and specificity of ERUS, CT, and MRI were respectively: (93%,74%), (79%,75%), and (90%,74%) for T-category and (75%,72%), (51%,88%), (60%,82%) for lymph nodes. Subgroup analysis including 169 patients from six studies was performed for MRI with

endorectal coil (which represents different technology and thus a different index test) and resulted in 89% sensitivity and 79% specificity for T-category and 82% sensitivity and 83% specificity for lymph nodes (no separate analysis by RT status was performed in this subgroup). The authors concluded that MRI with endorectal coil was the most accurate modality for both T-category and lymph nodes.

The same comparison between ERUS, CT, and MRI was repeated in a second systematic review by Bipat et al (53) based on a more comprehensive search of Medline, EMBASE, Cochrane, and CANCERLIT databases. The analysis was based on 90 studies (published 1985-2002), 45 of which overlapped with the analysis by Kwok. Similar to the first analysis, multiple MRI coil types were included; however, this study excluded patients who received preoperative RT or CRT. A bivariate random effects model (see Chapter 3.3) was used to compare summary estimates of sensitivity and specificity of each imaging modality for T-category and lymph node metastases. Diagnostic accuracy of T-category was separately assessed for each of muscularis propria invasion (T2 or higher versus T1), perirectal tissue invasion (T3 or higher versus T2 or lower), and adjacent tissue invasion (T4 versus T3 or lower). In the assessment of perirectal tissue invasion (T-category), the sensitivity of ERUS (90%, 95%CI 88-92) was significantly higher than both CT (79%, 95%CI 74-84) and MRI (82%, 95%CI 74-87), but the specificity of the three modalities was similar: (75%, 95%CI 69-81), (78%, 95%CI 73-83), and (76%, 95%CI 65-84) respectively. ERUS, CT, and MRI were also similar in their sensitivity (67%, 95%CI 60-73; 74%, 95%CI 67-80; 66%, 95%CI 54-76, respectively) and specificity (78%, 95%CI 71-84; 55% , 95%CI 43-67; 76%, 95%CI 59-87, respectively) to detect lymph node metastases. ERUS was more specific (86%, 95%CI 80-90) than MRI (69%, 95%CI 52-82) in distinguishing T2 or higher versus T1 tumours, but sensitivity of the two modalities was similar for this assessment. Sensitivity and specificity were also not significantly different between ERUS, CT, and MRI for detection of adjacent organ invasion.

The accuracy of ERUS, CT, and MRI for detecting lymph node metastases was further reviewed in a third meta-analysis (63) which also examined the accuracy of MRI for predicting CRM involvement. The review included 75 and 7 articles (published 1985-2005) examining lymph node metastases and CRM involvement, respectively. As in the meta-analysis by Kwok, patients who received preoperative RT or CRT were included. The type of MRI coil used in each study

was not described in the analysis. For each imaging modality, summary receiver operating characteristics (sROC) curves were constructed for each outcome using the Moses-Littenberg method, and diagnostic odds ratios (DOR) (see Chapter 3.3) were also calculated for detecting lymph node metastases. All three modalities performed poorly for detecting lymph node metastases (DOR 8.83 for EUS, 5.86 for CT, and 6.53 for MRI) with no significant difference between them. However, MRI had good accuracy for CRM assessment, with a sensitivity of 80% associated with a false positive rate of 20%. Separate subgroup analysis was not performed for patients who received preoperative RT or CRT.

Most recently, a systematic review by Purkayastha et al (64) evaluated MRI for CRM assessment in nine studies published 2001-2005. Patients who received preoperative RT or CRT were included in the analysis. A fixed effects model was used to estimate sensitivity and specificity with 95% confidence intervals, and a sROC curve was generated and used to estimate DOR. Study quality was formally assessed using a validated checklist. The overall sensitivity, specificity, and DOR of MRI for detecting CRM involvement were 94% (95%CI 90-97), 85% (95%CI 81-89), and 57.21 (95%CI 18-180), respectively. Subgroup analysis including five studies that used phased array coil produced sensitivity, specificity, and DOR estimates of 97% (95%CI 93-99), 84% (95%CI 75-91), and 68 (95%CI 13-350).

While these meta-analyses have provided some insight into the ability of MRI to stage rectal cancer, they suffer from a number of important limitations. First, none of the meta-analyses have examined all three of the elements important to local staging of rectal cancer (T-category, lymph nodes, CRM), and it is difficult to compare accuracy estimates derived from different meta-analyses using different inclusion criteria and methodology. Second, the larger and more comprehensive reviews (52,53) examined mainly older studies (published before 2000), many of which used MRI with body coil. This type of coil is outdated and no longer considered adequate for rectal cancer staging, with phased array coil being the current standard used in practice. These analyses have therefore included outdated technology and their results cannot be generalized to MRI technology as it currently exists. Third, inclusion criteria were broad in all four studies, resulting in important clinical heterogeneity in the technology and populations examined, but this heterogeneity was poorly managed. Two studies combined different types of technology (endorectal coil, phased array coil) without investigating the effect of individual coil

type on results (53,63) while two (63,64) of the three reviews (52,63,64) that included patients who underwent preoperative RT or CRT did not produce separate results by therapy status. Furthermore, three of the analyses used simple, fixed-effects statistical models which did not account for the correlation between sensitivity and specificity (52) or failed to account for heterogeneity between studies (47,63,64). Such models are not recommended for meta-analyses of diagnostic accuracy (62) (also see Section 3.3). Finally, three of the reviews (52,53,63) did not perform a formal assessment of study methodological quality, which is a critical component of meta-analyses involving observational studies (65).

2.2 Objective

The objective of this thesis was to perform a systematic review and meta-analysis of the literature to determine the accuracy of phased array pelvic MRI for 1) T-category, 2) lymph node metastases, and 3) CRM involvement.

Chapter 3
Methods

3.1 Study selection and data extraction

A comprehensive search of Medline (January 2000-March 2011), EMBASE (January 2000-March 2011), the Cochrane Central Register of Controlled Trials (through first quarter 2011), and the Cochrane Database of Systematic Reviews (2005-March 2011) was performed using search terms related to MRI, rectal cancer, neoplasm staging, and diagnostic assessment (Figure A.1). Reference lists of articles retrieved were also searched manually to identify relevant studies.

Studies were included in the systematic review if they fulfilled the following *a priori* defined criteria: 1) original published reports with primary data, 2) phased array pelvic coil MRI used to assess the accuracy of T-category, lymph nodes, or CRM involvement for biopsy-proven primary rectal cancer, 3) histopathology used as the reference standard, and 4) English language. To ensure that technology under assessment was relatively current, final results were restricted to studies published since January 2000. Studies in which original primary data were available to extract or reconstruct 2x2 contingency tables for the outcomes of interest were eligible for inclusion in the meta-analysis. Where raw data were not published but an article was otherwise eligible for inclusion, authors were contacted to request this information. Excluded from the meta-analysis were: 1) studies that incorporated other MRI coil types (e.g., endorectal or body coil) which were deemed to represent a fundamentally different index test and are less widely used in practice, 2) duplicate data from another included study, and 3) patients who received preoperative CRT or long course RT alone. These patients were excluded because both long course RT and CRT can cause tumour regression, resulting in a final histopathologic stage that is less advanced (downstaged) compared with the tumour at initial imaging. In such patients, histopathology would no longer be an acceptable reference standard for MRI. In cases where only a portion of the study population met the inclusion criteria (e.g., some patients received preoperative RT and some did not), only data for those eligible patients were included in the analysis if they could be extracted separately.

Titles and abstracts identified by the search strategy were independently reviewed by three reviewers. For all abstracts that met inclusion criteria or were potentially eligible, full articles were retrieved and independently reviewed by two reviewers.

Data from selected articles were independently abstracted by two reviewers using a standardized data extraction form and included study variables (year of publication, study design), test interpretation variables (definitions for MRI "positive cases", number of readers, approach to image assessment), population characteristics (inclusion and exclusion criteria), MRI protocol (field strength, sequences, resolution parameters, contrast use), and test results (2x2 contingency tables). Each paper was then discussed among all reviewers and discrepancies were resolved by consensus. In studies where separate results were provided for multiple independent readers, each set of results was included as a separate dataset.

Two-by-two contingency tables were extracted or reconstructed for each outcome on a patient level as follows: a) T-category - T3/T4 versus T1/T2, b) Lymph nodes - any involved versus uninvolved, c) CRM - involved versus uninvolved. This categorization was based on criteria from clinical guidelines for preoperative CRT (55), in order to determine how well MRI classifies patients who should or should not receive this treatment. Definitions of positive cases on MRI differed between studies and were extracted for each study when available.

3.2 Assessment of methodological quality

A modified version of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was used to assess the methodological quality of studies included in the meta-analysis. QUADAS is a 14-item checklist developed specifically to assess the quality of primary studies for inclusion in systematic reviews and meta-analyses (66). It has been adopted by the Cochrane Collaboration (65) as a guide to the conduct and reporting of diagnostic accuracy studies. The development of the tool followed a rigorous and methodologically accepted approach suggested by Norman and Streiner (67) and adopted by Jadad et al (68) to develop a scale to assess the quality of randomized controlled trials. Evidence from the published literature and previously developed tools was comprehensively synthesized and considered by methodology experts using a Delphi procedure to produce the final tool.

QUADAS items fall within three domains: bias (9 items), generalizability (2 items), and quality of reporting (3 items) (Table A.4). Each item is scored as yes/no/unclear. Bias items include elements of study design and execution (items 3,4,5,6,7,12,14), and interpretation (items 10,11), and generalizability items refer to the composition of patients selected (items 1,2). The Cochrane Diagnostic Test Accuracy (DTA) Handbook recommends using a modified version of QUADAS which omits three items that focus on quality of reporting rather than methodological quality (items 2,8,9) (65). It is also recommended that items important to a particular topic be added to the checklist as necessary. In this review, an important determinant of study quality was whether the index test (MRI) and reference standard (histopathologic assessment) were appropriately performed. Therefore, the following two items were added to the QUADAS checklist: 1) Does the execution of the index test meet accepted standards? and 2) Does the execution of the reference standard meet accepted standards? The modified version of QUADAS used in this review is attached in Table A.5.

As a short, simple checklist developed specifically for use in DTA meta-analyses, QUADAS is the best available tool to assess the methodological quality of studies in this review. It does, however, have some important limitations. First, the tool was created with a generic purpose and intended to apply to all diagnostic accuracy studies rather than studies of radiological imaging in particular. Because of this, items are broadly defined and vaguely phrased, leaving considerable room for interpretation. For example, item 2, "Were selection criteria clearly described?" can be interpreted to refer to age, sex, disease stage, tumour location, co-morbidities, or other selected variables. Minimum criteria for fulfilling each QUADAS item must therefore be determined by reviewers before the tool can be applied. Second, the inter-observer reliability of the tool is variable, with agreement on individual items ranging from "no agreement beyond chance" to "almost perfect agreement" (69-71) based on calculated kappa values. Third, QUADAS is not intended by its developers to produce a total score because i) the weighting and calculation of quality scores are felt to be arbitrary and subjective, and ii) quality scores ignore the variable importance of individual items in different contexts (66). Instead, the authors recommend individual assessment of QUADAS items to determine risk of bias in each study. The lack of an aggregate score limits the utility of the checklist within a meta-analysis, as overall study quality cannot be directly incorporated into the results.

In applying QUADAS to this meta-analysis, minimum criteria for fulfilling each item were discussed by reviewers and disagreements were resolved by consensus prior to scoring studies. In keeping with Cochrane DTA Working Group recommendations, no summary score was calculated.

3.3 Statistical analysis and data synthesis

3.3.1   Measures of diagnostic accuracy

Common indices of diagnostic accuracy include sensitivity and specificity, positive and negative likelihood ratios, positive and negative predictive values, and the diagnostic odds ratio. Sensitivity is the proportion of patients with the disease who are correctly identified to have it (true positive rate), while specificity is the proportion of patients without disease who are correctly identified to be free of it (true negative rate). Predictive values are related to disease prevalence within a population and are calculated based on index test results as the proportion of test positive cases that are true positives (PPV) or the proportion of test negative cases that are true negatives (NPV). Likelihood ratios are calculated as functions of sensitivity and specificity. A positive likelihood ratio (LR+ = sensitivity/[1-specificity]) estimates how much higher the odds are that a positive index test result would be found in a diseased versus a non-diseased patient, while a negative likelihood ratio (LR- = [1-sensitivity]/specificity) estimates how much lower the odds are that a negative index test result would be found in a diseased versus a non-diseased patient. The diagnostic odds ratio (DOR) is the ratio of odds of a positive test in patients with disease relative to patients without disease (i.e., LR+/LR-).

The choice of summary index to be used in a meta-analysis can be guided by its properties. Sensitivity and specificity are familiar to clinicians and are (theoretically) independent of disease prevalence, making them a good choice to summarize data from multiple studies with heterogeneous populations. They can also be used to calculate likelihood ratios, if desired. Sensitivity and specificity provide different information about a given test. Comparing tests with a focus on sensitivity is useful when the objective is to identify the test that is least likely to miss a positive case (or understage), while comparing tests with a focus on specificity is useful when the objective is to identify the test that is least likely to falsely diagnose a negative case as positive (or overstage). When the objective is to identify the test with the maximum "overall"

diagnostic performance, DOR is an appropriate measure. As a function of sensitivity and specificity, DOR is similarly (theoretically) independent of disease prevalence and is useful as a single indicator of test performance that can allow comparison between different tests. Higher DOR values indicate better test performance (72).

In reality, it has been demonstrated that sensitivity and specificity (and by extension, likelihood ratios and DOR) can actually vary with prevalence. One reason for this is that the distinction between the presence or absence of disease may not be truly dichotomous but rather may lie along a spectrum of disease severity from which a particular threshold is used to identify positive cases. The distribution of disease severity in a test population relative to the selected threshold affects not only the prevalence but also the likelihood of misclassification (i.e., cases which lie closer to the threshold are more likely to be misclassified than cases which lie at the extremes of the severity spectrum). Thus, the sensitivity and specificity of the test will vary with prevalence, a phenomenon known as "spectrum effect" (73). Furthermore, where test results require human interpretation (e.g., a radiologist reviewing MRI scans), the clinician interpreting the results of a diagnostic test may be aware of the prevalence of the disease in question in the population under assessment and may be more likely to classify a given result as positive in a high prevalence population (and vice versa in a low prevalence population). Therefore, the same test would have high sensitivity but low specificity in high prevalence populations, and low sensitivity but high specificity in low prevalence populations (74). Despite these limitations, sensitivity, specificity, and DOR are appropriate summary measures to use in meta-analysis provided that care is taken that included studies have sufficiently similar distribution of disease severity and prevalence.

3.3.2   Approach to diagnostic accuracy data

Traditional meta-analytic methods were developed in order to synthesize the results of intervention studies. These methods essentially involve pooling the treatment effect measures derived from individual studies to calculate a weighted average of intervention effects (e.g., odds ratios for dichotomous outcomes). If studies are sufficiently homogeneous in population and methodology, a meta-analysis may use a fixed effect statistical model which assumes that the true effect size is the same across studies. Fixed effects models account for within study variability but do not allow for between study variability. Each model typically deals with only one primary outcome (75). In contrast, meta-analyses of diagnostic test accuracy (DTA) studies

must simultaneously deal with two correlated summary statistics (sensitivity and specificity) and must account for the impact of the positivity threshold used across different studies on this relationship. Using the same index test, a higher cut-off to define positive cases would result in higher specificity and lower sensitivity, while a lower cut-off would result in higher sensitivity and lower specificity. The values of sensitivity-specificity pairs at each threshold can be plotted in receiver operating characteristics (ROC) space to demonstrate this correlation for a given test. Threshold differences can arise not only due to explicit differences in positivity cut-off values between studies (e.g., what distance between tumour and mesorectal fascia constitutes a positive CRM) but also due to implicit variations in how these values are measured or interpreted (e.g. where to measure this distance on the MRI image). Threshold effects are particularly important in tests which utilize a cutoff along a continuous or ordinal scale to define test positivity; however, they may still be relevant in tests where interpretation of results as positive or negative is subjective, since the criteria for determining positivity in such tests may lie along an implicit scale (e.g., degree of signal heterogeneity indicating lymph node involvement) such that sensitivity and specificity vary in an ordered fashion depending on the particular cutoff chosen to define a positive test.

Statistical tests such as the I^2 statistic, which can be used to check for heterogeneity in meta-analyses of intervention studies, are not useful in DTA meta-analyses because they do not account for heterogeneity due to threshold effects. As a result, the Cochrane Collaboration experts recommend assuming that heterogeneity is nearly always present to some degree in DTA meta-analyses and must be considered when selecting a statistical model (62).

3.3.3 Meta-analytic models

The simplest statistical approach, involving separate pooling of sensitivity and specificity or other accuracy indices does not account for correlation between these measures and gives average estimates of these values at some unspecified mean threshold (which may not even be clinically applicable) (76). A more advanced approach proposed by Moses and Littenberg (77,78) involves estimating a summary receiver operating characteristics (sROC) curve to account for variations in thresholds between studies. In this model, the pairs of sensitivity (true positive rate or TPR) and 1-specificity (false positive rate or FPR) for the ith study are logit-transformed to produce two new variables, D_i and S_i as follows:

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

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

where D_i is the log of the diagnostic odds ratio and S_i represents a measure of the threshold used. The two new variables are then used to fit a linear model with S_i as the independent variable and D_i as the dependent variable. In the final step of the model, the pairs of D_i and S_i are anti-logit transformed into ROC space to produce a summary curve for each outcome. The position of the summary ROC curve indicates the difference in diagnostic performance of a test; the closer a curve is located to the upper left corner of the ROC space, the better the diagnostic performance.

While the Moses-Littenberg method has been widely applied in DTA meta-analyses, it is a fixed effects model which does not account for random variation between studies and ignores correlations in data within studies (79). Although it enables estimation of sensitivity for a given value of specificity, it does not produce average values of sensitivity and specificity (80). Furthermore, the properties of its slope (shape) parameter mean that in some datasets (for instance, where there is heterogeneity in test accuracy between subgroups of studies), improper sROC curves can result in which both sensitivity and specificity decrease together (which does not make sense clinically). Another limitation of this model is that it does not take into account the sampling error of S_i, treating it as a known parameter when it is in fact derived from TP and FP. (62)

The limitations of the Moses-Littenberg model have led to the development of newer, more complex statistical models. Two of these are advocated for use by the Cochrane Collaboration: the bivariate model (81) and the hierarchical summary receiver operating characteristics (HSROC) model (82). Both are random effects approaches which offer the important advantage of being able to account for heterogeneity both within study samples (sampling variability) and between study samples (differences in population, readers, diagnostic technology, etc). Both models also allow for the addition of covariates to explore the influence of individual sources of heterogeneity on estimates. This offers an analytical advantage in that all studies can be included in the analysis to maintain statistical power, in contrast with traditional subgroup analysis where different subgroups are analyzed separately.

3.3.4   Bivariate model

In the first level of this generalized linear mixed effects model, within-study variability for sensitivity and specificity is assumed to follow a binomial distribution and sensitivity-specificity pairs from each study are jointly modeled using logit-transformations. This level accounts for heterogeneity within studies arising from sampling variability. For sensitivity (*A*) in a given study *i*, the number of patients who have a positive test is specified by:

$$y_{Ai} \sim \text{Binomial}(n_{Ai}, \text{Sens}_i), \text{logit}(\text{Sens}_i) = \beta_{Ai} + \mu_{Ai}$$

where n_{Ai} represents the total number of patients with disease, Sens_i represents the probability of a positive test result in that group (i.e., sensitivity) in study *i*, and β_{Ai} represents covariates that may be related to sensitivity. Similarly, for specificity (B) in a given study *i*, the number of patients who have a negative test is specified by:

$$y_{Bi} \sim \text{Binomial}(n_{Bi}, \text{Spec}_i), \text{logit}(\text{Spec}_i) = \beta_{Bi} + \mu_{Bi}$$

where n_{Bi} represents the total number of patients without disease, Spec_i represents the probability of a negative test result in that group (i.e., specificity) in study *i*, and β_{Bi} represents covariates that may be related to specificity.

In the second level, the logit-transformed sensitivity and specificity are modeled using a single bivariate normal distribution to account for the correlation between sensitivity and specificity. Here, assumptions of random effects take into account heterogeneity arising from true differences in sensitivity and specificity across studies. The second level of the model takes the form:

$$\begin{bmatrix} \mu_{Ai} \\ \mu_{Bi} \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_{Ai} \\ \mu_{Bi} \end{bmatrix}, \Sigma \right) \quad \text{with } \Sigma = \begin{bmatrix} \sigma^2_A & \sigma_{AB} \\ \sigma_{AB} & \sigma^2_B \end{bmatrix}$$

where μ_A is mean logit sensitivity, μ_B is mean logit specificity, σ^2_A is the variance of logit sensitivity, σ^2_B is the variance of logit specificity, and σ_{AB} is the covariance between logit sensitivity and specificity. Associated 95% confidence and prediction regions around sensitivity-specificity pair estimates (depicted as ellipses around a summary point on a sROC plot) can also be calculated. The prediction ellipse demonstrates the extent of heterogeneity in the data as it

indicates the region within which one can predict with 95% confidence that the "true" sensitivity and specificity of a future study should be (83).

3.3.5   HSROC model

Similar to the Moses-Littenberg model, the HSROC approach models functions of sensitivity and specificity to estimate accuracy and threshold parameters for a summary ROC curve but is designed to overcome the limitations of the former approach. Like the bivariate model, it involves two levels, logit-transforming TPR and FPR pairs jointly using a binomial distribution to account for within-study variability at the lower level and fitting the transformed variables as a random effects model with assumed normally distributed variability to account for between-study heterogeneity at the higher level. In the first level, the number of patients testing positive in the diseased and non-diseased groups in a given study i are assumed to follow binomial distributions as given by $y_{ij} \sim \text{Binomial}(n_{ij}, \pi_{ij})$, where j denotes disease group (1=diseased, 2=non-diseased), n_{ij} represents the total number of patients tested and π_{ij} represents the probability of a positive test result. The model is written as:

$$\text{logit}(\pi_{ij}) = (\theta_i + \alpha_i X_{ij})\exp(-\beta X_{ij})$$

where θ_i is a proxy for the positivity threshold (calculated as the mean of the log odds of a positive test result for the diseased and the log odds of a positive test result for the non-diseased groups), α_i is the lnDOR, and X_{ij} represents the true disease status for patients in the given disease group. β is a scale parameter which describes the shape of the sROC curve. Since the sROC curve is constructed from data points derived from multiple studies, β cannot be estimated based on any single study and is therefore assumed to be the same across all studies (i.e., it is fitted as a fixed effects parameter) (62,82).

In the second level of the HSROC model, accuracy and threshold derived from each study are assumed to be independent and normally distributed with $\alpha i \sim N(\Lambda, \sigma_\alpha)$ and $\theta i \sim N(\Theta, \sigma_\beta)$, where Λ is mean accuracy, σ_α is variance of accuracy, Θ is mean threshold, and σ_β is variance of threshold. To construct a sROC curve, the mean accuracy and shape parameters estimated by the model can be used to calculate sensitivity (TPR) from a selected range of FPR values as follows:

$$\text{Sensitivity} = 1/[1+\exp(-(\Lambda e^{-0.5\beta}+\text{logit(FPR)}e^{-\beta}))]$$

3.3.6   Considerations for selecting a model

In the absence of additional covariates, the bivariate and HSROC models have been shown to be identical (83) and simply offer different perspectives on the same data through specification of different parameters to provide a mean estimate at a selected threshold (bivariate model) or a summary curve encompassing all possible thresholds (HSROC model).   The models remain equivalent if the same covariates are added to each and allowed to affect both accuracy and threshold parameters (HSROC model) or both sensitivity and specificity (bivariate model) (83). The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy recommends using HSROC modeling to estimate a sROC curve when thresholds are highly variable between studies (e.g., when various distances are used to define CRM involvement), in order to demonstrate the range of clinically possible sensitivity-specificity values, and using bivariate modeling to estimate an average sensitivity and specificity at a given threshold when this cutpoint is common across studies (62).   In cases where it is unclear what threshold has been used in different studies or this cutpoint is subjective (e.g., the criteria for mesorectal fat invasion in T3 tumours), both models may be used to provide complementary information, as has been done in the meta-analysis reported here.

3.3.7   Applying the model

In this meta-analysis, statistical analysis was performed using bivariate random effects modeling and hierarchical summary receiver operating characteristics (HSROC) modeling.   Using SAS software (SAS institute, Cary, NC), bivariate random effects modeling was used to produce summary estimates of sensitivity and specificity while HSROC modeling was used to estimate the diagnostic odds ratio (DOR) for each outcome with 95% confidence intervals for each estimate.   Sensitivity, specificity, and DOR were compared between outcomes using the z-test for equality of proportions, and significance was set at $p<0.05$.   Once parameter estimates (means and variances) were obtained, they were imported into the Cochrane Collaboration's Review Manager software (RevMan 5).   This program uses the formulae of the bivariate and HSROC models as described above to fit summary ROC curves with 95% confidence and prediction regions.

3.4 Investigations of heterogeneity

Covariate analyses were planned for variables which were determined to be potential sources of heterogeneity in MRI diagnostic performance. The following covariate analyses were planned for each outcome: MRI protocol (adequate versus inadequate), field strength (3T versus lower fields), definition of T3 tumour (nodular versus spiculated tumour margin), definition of lymph node metastatic involvement (size versus border versus signal criteria), definition of CRM involvement (<1mm versus <2mm from mesorectal fascia), approach to image assessment (independent versus consensus). MRI protocols were deemed adequate if they included T2 weighted, high resolution axial images performed perpendicular to long axis of tumour-bearing rectum. Investigating one variable at a time, each covariate was allowed to influence sensitivity and specificity (and by extension, DOR) for each outcome and a covariate was considered statistically significant if it altered an estimate with a p-value of $p<0.05$ (likelihood ratio test). Sensitivity analyses were also planned to investigate the influence of individual quality items on results.

Chapter 4
Results

4.1 Search results and study selection

The initial literature search identified 1145 abstracts of which 95 were judged relevant for full assessment; an additional 9 articles were identified from the reference lists of these articles and were also read in full. Following data extraction, 21 articles (84-104) met inclusion criteria and have been included in this meta-analysis. Requests for original data from authors of 14 potentially eligible articles yielded no further information, and these were therefore excluded. The remaining articles did not meet inclusion criteria for a number of reasons and were excluded (FIGURE 4.1).

4.2 Search characteristics

Data were extracted from 19 studies for T-category assessment in 1986 patients (33 datasets), from 13 studies for lymph node assessment in 1249 patients (23 datasets), and from 10 studies for CRM assessment in 986 patients (13 datasets). Characteristics of included studies are summarized in TABLE 4.1. Study design was retrospective in seven studies, prospective in 12 studies, and unclear in two. Six studies employed a consensus of two radiologists and 14 recorded independent readings by one or more radiologists; this information was not available in one study. There was notable heterogeneity among studies in the criteria used to define "positive" cases, and this information was frequently not provided. This made it difficult to interpret how the mean estimates of sensitivity and specificity derived for each staging element might apply to any particular definition of positivity for that element. In the assessment of T-category, "spiculated" extensions from the tumour into the perirectal fat were considered T3 tumour in some studies and T2 tumour in others. Lymph node size $\geq$5mm, irregular borders, and mixed signal intensity were used in different combinations across studies to define lymph node metastatic involvement. CRM involvement was most variably defined, ranging from direct contiguity of tumour with the mesorectal fascia to tumour extension $\leq$5mm from the mesorectal fascia.

Figure 4.1 Flow diagram for identification of eligible studies

Titles and abstracts
N = 1145

1050 discarded because not
original reports or not about MRI
accuracy for rectal cancer

Assessment of full
text
N = 95

Articles identified from reference
lists
N = 9

Total full text
assessed
N = 104

83 excluded from the review:
- examined other outcomes (25)
- reference standard not
histopathology (3)
- used different coil type (16)
- included non-rectal cancer or
recurrent rectal cancer patients
(4)
- could not extract raw data (14)
- examined only patients who
received nCRT or separate data
not available for surgery only
group (18)
- duplicate data from another
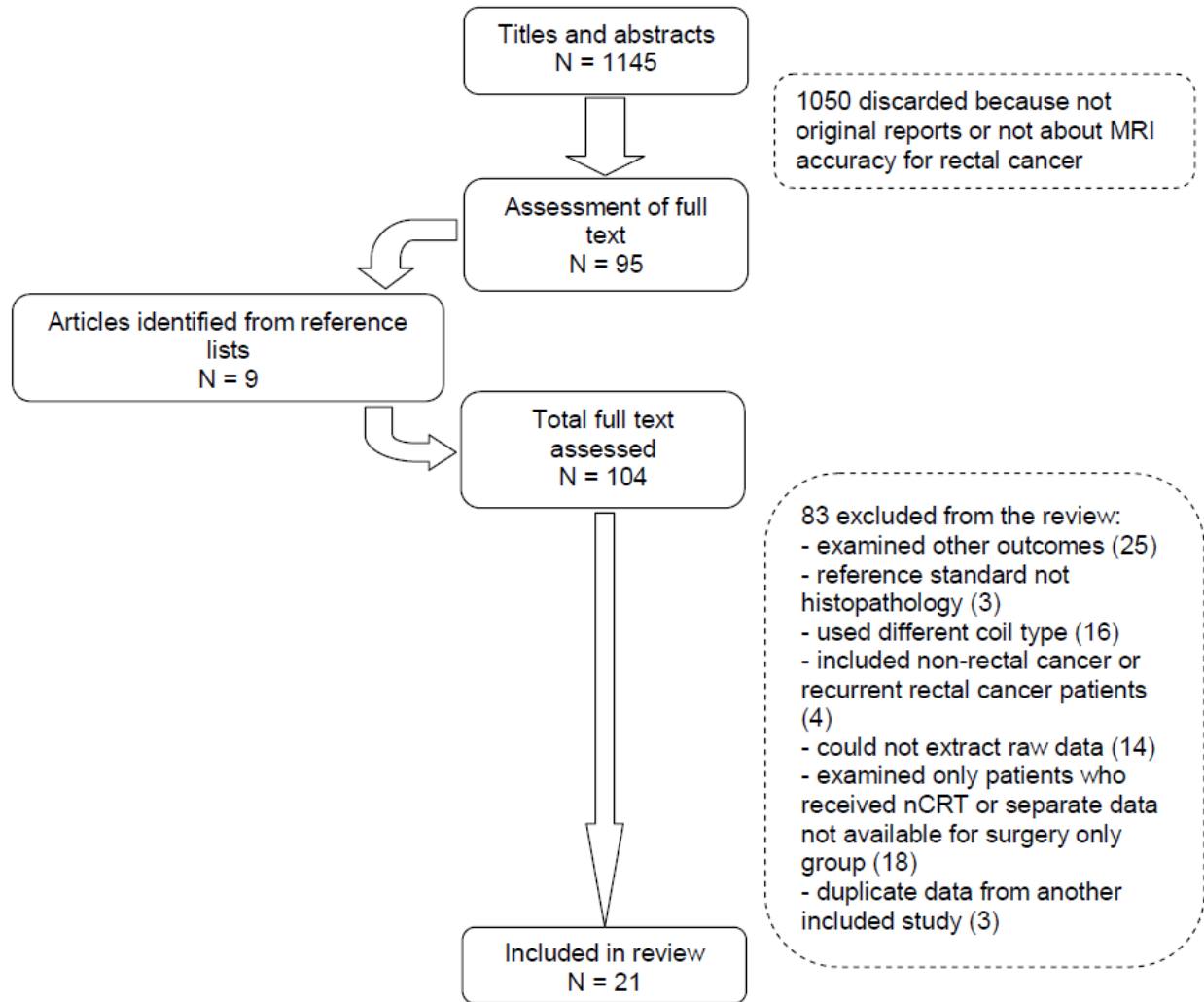included study (3)

Included in review
N = 21

Table 4.1 Characteristics of studies included in the analysis

Study	Design	Assessment approach (# of readers for each MRI)	Elements assessed	# of patients in analysis	Definition of positive case
Gagliardi 2002 (87)	prospective	independent (1)	T-category	28	T3 = spiculation into perirectal fat
			Nodes	28	Node diameter >5mm
Branagan 2004 (84)	prospective	independent (1)	T-category	40	Not provided
			Nodes	40	Not provided
			CRM	40	Contiguity of tumour with MRF
Kim 2004 (89)	prospective	independent (4)	T-category	62	T3 = spiculation into perirectal fat
			Nodes	62	Node diameter >5mm or heterogeneous signal or irregular border
Ferri 2005 (86)	unclear	consensus (2)	T-category	33	Not provided
			Nodes	33	Node diameter $\geq$5mm
			CRM	33	Tumour extension within $\leq$1mm of MRF
Oberholzer 2005 (91)	prospective	consensus (2) [†]	T-category	19	Not provided
			Nodes	17	Node diameter >5mm with inhomogeneous decreased signal and irregular border
Vliegen 2005 (96)	retrospective	independent (2)	T-category	56	T3 = spiculations in the perirectal fat or a bulging tumour margin beyond the contour of the rectal wall
			CRM	56	Clear spiculations from tumour reaching MRF or minimal fat plane between tumour/tumour deposit/suspected node and MRF
Burton 2006 (85)	retrospective	independent (1)	T-category	57	Not provided
			Nodes	57	Not provided
			CRM	57	Not provided
Kim 2006 (99)	prospective	independent (3)	T-category	35	Not provided
			Nodes	30	Indistinct border or irregular margin or mixed signal intensity
MERCURY 2006 (90)	prospective	independent (1)*	CRM	311	Tumour extension within $\leq$1mm of MRF
MERCURY 2007 (104)	prospective	independent (1)*	T-category	300	Not provided
Rao 2007 (93)	prospective	consensus (2)	T-category	67	T3 = spiculation into perirectal fat
			CRM	67	Tumour extension within 2mm of MRF
Strassburg 2007 (94)	retrospective	independent (1)	T-category	43	Not provided
			Nodes	43	Not provided
			CRM	45	Tumour/tumour deposit/involved node/EMVI within $\leq$1mm of MRF
Taylor 2007 (95)	retrospective	independent (1)	CRM	42	Tumour extension within $\leq$5mm of MRF
Futterer 2008 (103)	prospective	consensus (2)	T-category	22	T3 = irregular spiculation into perirectal fat

Table 4.1 (continued)

Study	Design	Assessment approach (# of readers for each MRI)	Elements assessed	# of patients in analysis	Definition of positive case
Halefoglu 2008 (88)	unclear	Independent (1)	T-category	34	T3 = spiculation into perirectal fat
			Nodes	34	Node diameter $\geq$5mm
Kim 2008 (102)	retrospective	independent (2)	T-category	42	Thickened but not fine spiculations
Piippo 2008 (92)	prospective	consensus (2)	T category	26	T3 = gross spiculations into perirectal fat
Akasu 2009 (97)	prospective	consensus (2)[†]	T-category	101	T3 = nodular projection into perirectal fat
			Nodes	104	Node diameter $\geq$5mm
			CRM	101	Not provided
Kim 2009 (98)	prospective	independent (1)	T-category	66	Not provided
			Nodes	66	Irregular border or mixed signal intensity
			CRM	66	Tumour <1mm from MRF
Kam 2010 (100)	retrospective		T-category	23	Not provided
			Nodes	23	Not provided
Kim 2010 (101)	retrospective	independent (2)	T-category	109	Not provided
			Nodes	109	Indistinct border, irregular margin, or mixed signal intensity

CRM, circumferential resection margin; EMVI, extramural vascular invasion; MRF, mesorectal fascia; MRI, magnetic resonance imaging
[†] One assessor was a surgeon
*Multicentre study; 18 different radiologists, each MRI only read by one radiologist

4.3 Methodological quality

The methodological quality of included studies is summarized in TABLE 4.2. As per the inclusion criteria, all studies used an acceptable reference standard (histopathology). In general, poor reporting of study details prohibited assessment of studies on a number of quality items, particularly those related to patient spectrum, execution of the MRI and histopathological assessment. This made it difficult to use the QUADAS tool to distinguish between high and low quality studies.

Table 4.2 Methodological quality of studies included in the final analysis

Study	Representative spectrum?	Acceptable reference standard?	Acceptable delay between tests?	Partial verification avoided?	Differential verification avoided?	Incorporation avoided?	Reference standard results blinded?	Index test results blinded?	Relevant clinical information?	Uninterpretable results reported?	Withdrawals explained?	Adequate MRI protocol?	Adequate histopathologic assessment?
Futterer 2008 (103)	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓
Rao 2007 (91)	?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kim 2006 (99)	?	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓
Kim 2010 (101)	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	?
Oberholzer 2005 (91)	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	?
Akasu 2009 (97)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	?	-
MERCURY 2006 (90)	✓	✓	?	✓	✓	✓	?	✓	✓	✓	✓	✓	?
MERCURY 2007 (104)	✓	✓	?	✓	✓	✓	?	✓	✓	✓	✓	✓	?
Piippo 2008 (92)	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	✓	?	?
Taylor 2007 (95)	✓	✓	?	✓	✓	✓	?	✓	✓	✓	✓	✓	?
Ferri 2005 (86)	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	?	?
Kim 2008 (102)	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	-	?
Kim 2009 (98)	-	✓	?	✓	✓	✓	?	✓	✓	✓	✓	✓	✓
Branagan 2004 (84)	✓	✓	?	✓	✓	✓	?	✓	✓	✓	✓	?	?
Vliegen 2005 (96)	?	✓	?	✓	✓	✓	?	✓	?	✓	✓	✓	✓
Kim 2004 (89)	-	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	?	?
Gagliardi 2002 (87)	?	✓	?	✓	✓	✓	?	✓	✓	✓	✓	?	?
Halefoglu 2008 (88)	✓	✓	?	✓	✓	✓	✓	?	?	✓	✓	?	?
Strassburg 2007 (94)	?	✓	?	✓	✓	✓	?	?	?	✓	✓	✓	✓
Burton 2006 (85)	-	✓	?	✓	✓	✓	?	✓	?	✓	✓	✓	?
Kam 2010 (100)	?	✓	?	✓	✓	✓	?	✓	?	✓	✓	-	✓

4.4 Data analysis

For each outcome (T-category, lymph node metastases, and CRM involvement), summary estimates of sensitivity, specificity, and DOR were produced with 95% confidence intervals (TABLE 4.3), and summary ROC curves were fitted with 95% confidence and prediction regions (FIGURES 4.2-4.4). The three elements were also plotted in a single ROC space for comparison (FIGURE 4.5). MRI specificity was significantly higher for CRM involvement (94%, 95%CI 88-97) than for T-category (75%, 95%CI 68-80; p<0.0001) and lymph node metastases (71%, 95%CI 59-81; p<0.0001). MRI sensitivity was significantly higher for T-category (87%, 95%CI 81-92) than for lymph node metastases (77%, 95%CI 69-84; p<0.0001) and CRM involvement (77%, 95%CI 57-90; p=0.004). DOR was significantly higher for CRM (56.1, 95%CI 15.3-205.8) than for lymph node metastases (8.3, 95%CI 4.6-14.7; p=0.0007) and, to a lesser extent, for T-category DOR (20.4, 95%CI 11.1-37.3; p=0.04).

Table 4.3 Sensitivity, specificity, and DOR of MRI for T-category, lymph node metastases and CRM involvement

Parameter	Sensitivity (%)	Specificity (%)	DOR
T-category	87 (95%CI 81-92)	75 (95%CI 68-80)	20.4 (95%CI 11.1-37.3)
Lymph node metastases	77 (95%CI 69-84)	71 (95%CI 59-81)	8.3 (95%CI 4.6-14.7)
CRM involvement	77 (95%CI 57-90)	94 (95%CI 88-97)	56.1 (95%CI 15.3-205.8)

CI, confidence interval; CRM, circumferential resection margin; DOR, diagnostic odds ratio

Figure 4.2 Estimated summary receiver operating characteristics (sROC) curve of MRI for assessment of T-category in rectal cancer. Individual squares represent the sensitivity-specificity pairs from individual studies in the analysis. Solid lines represent the range of possible sensitivity-specificity pairs based on all studies (from HSROC model). Solid circle represents the summary point estimates with associated 95% confidence region (fine dotted line) and 95% prediction region (coarse dotted line) (from bivariate model).

Figure 4.3 Estimated summary receiver operating characteristics (sROC) curve of MRI for assessment of lymph node metastases in rectal cancer. Individual squares represent the sensitivity-specificity pairs from individual studies in the analysis. Solid lines represent the range of possible sensitivity-specificity pairs based on all studies (from HSROC model). Solid circle represents the summary point estimates with associated 95% confidence region (fine dotted line) and 95% prediction region (coarse dotted line) (from bivariate model).
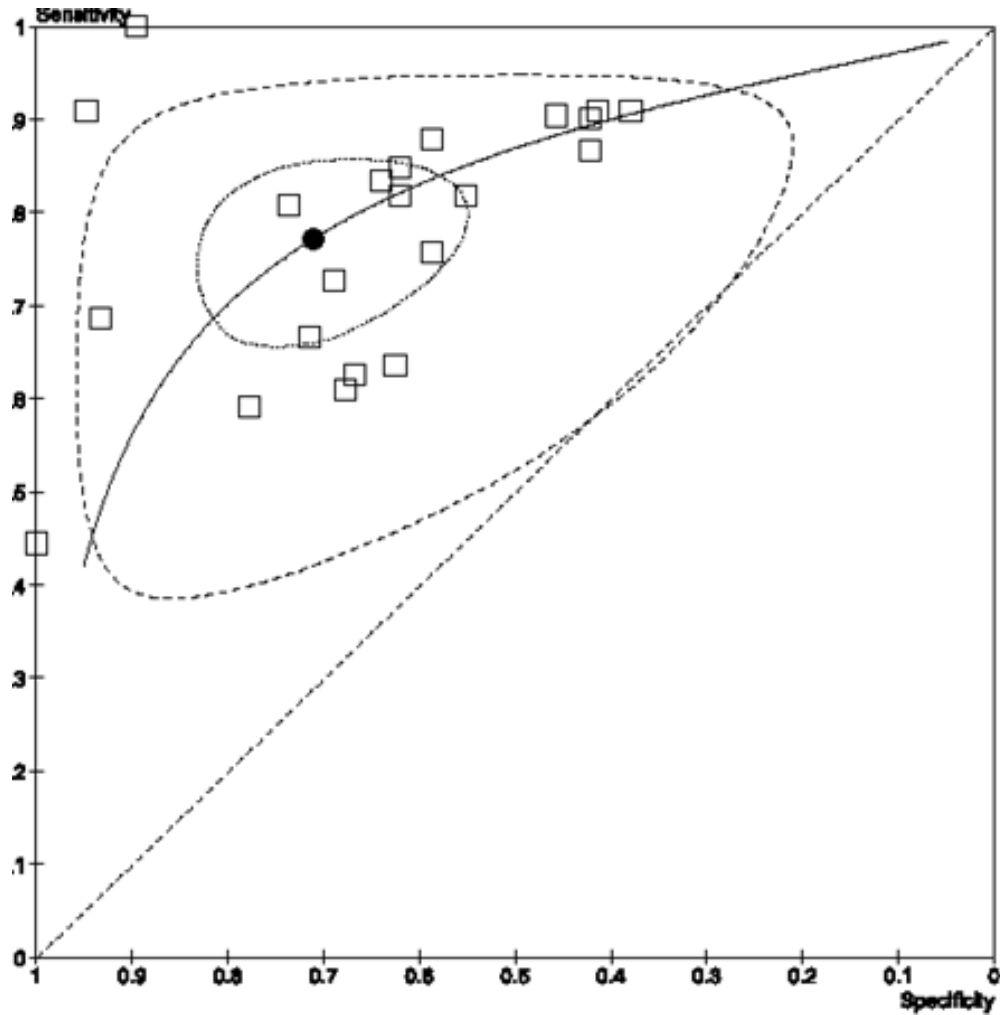
Figure 4.4 Estimated summary receiver operating characteristics (sROC) curve of MRI for assessment of circumferential resection margin (CRM) involvement in rectal cancer. Individual squares represent the sensitivity-specificity pairs from individual studies in the analysis. Solid lines represent the range of possible sensitivity-specificity pairs based on all studies (from HSROC model). Solid circle represents the summary point estimate with associated 95% confidence region (fine dotted line) and 95% prediction region (coarse dotted line) (from bivariate model).
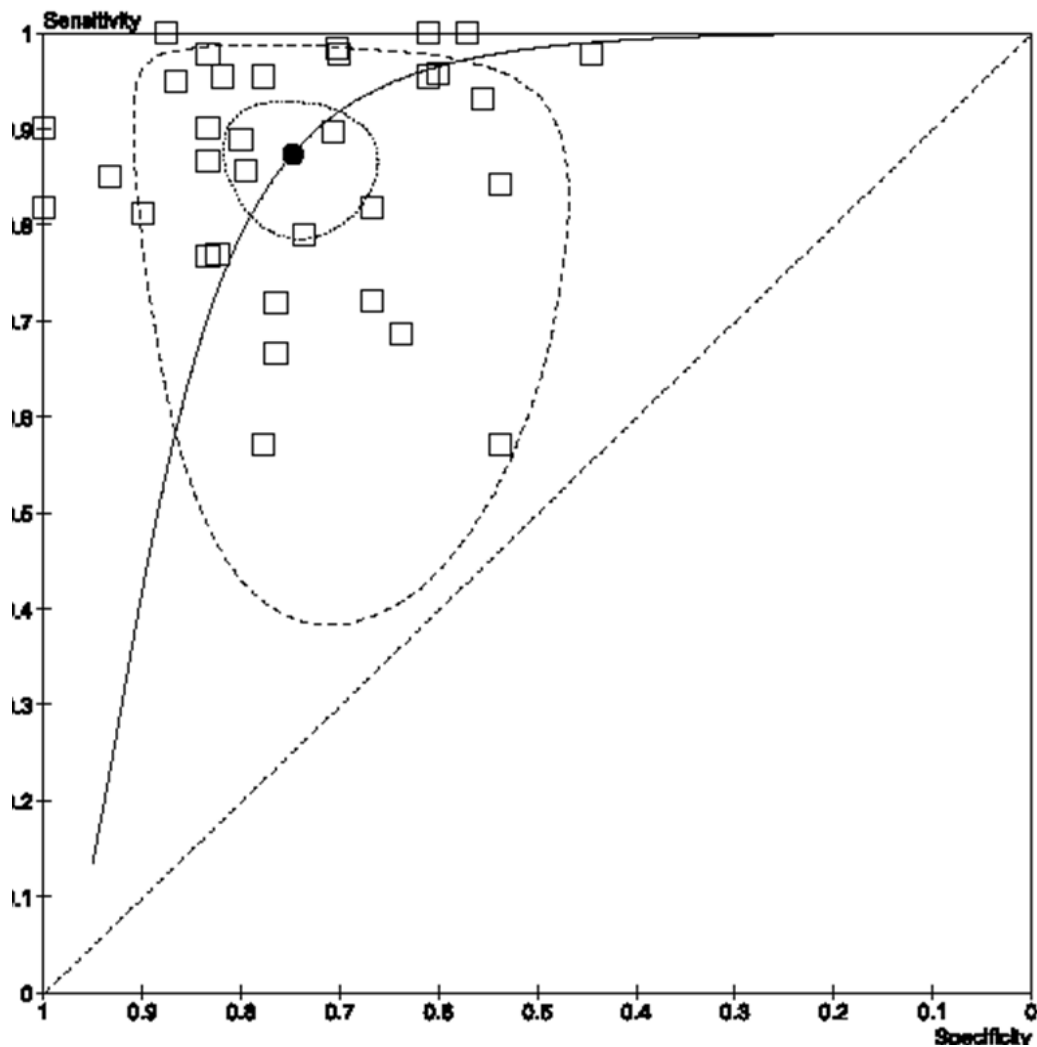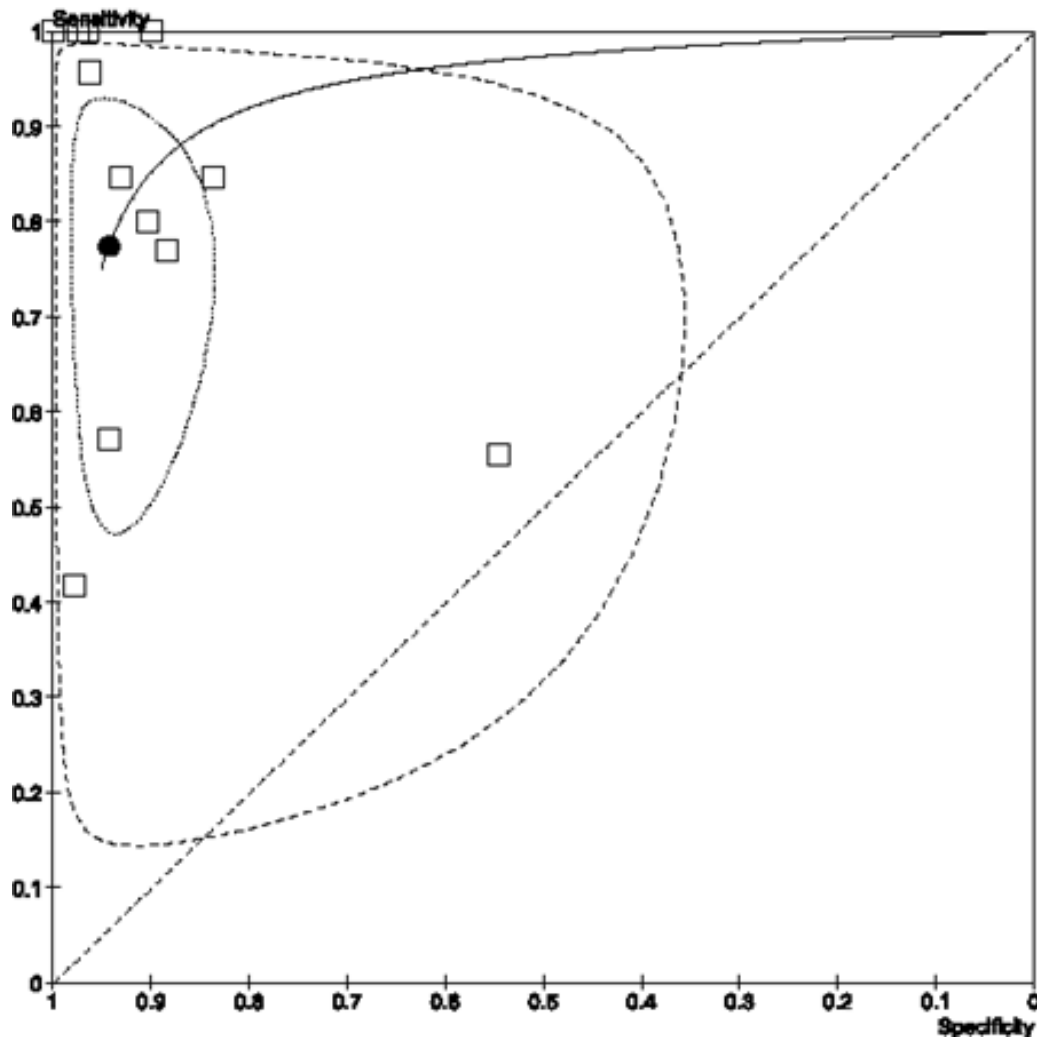
Figure 4.5 Estimated summary receiver operating characteristics (sROC) curve of MRI for assessment of T-category, mesorectal lymph nodes, and circumferential resection margin (CRM) involvement in rectal cancer. Solid lines represent the range of possible pairs of (sensitivity, specificity) for each staging element (from HSROC model). Solid circles represent the summary point estimates with associated 95% confidence regions (dotted lines) for each staging element (from bivariate model).
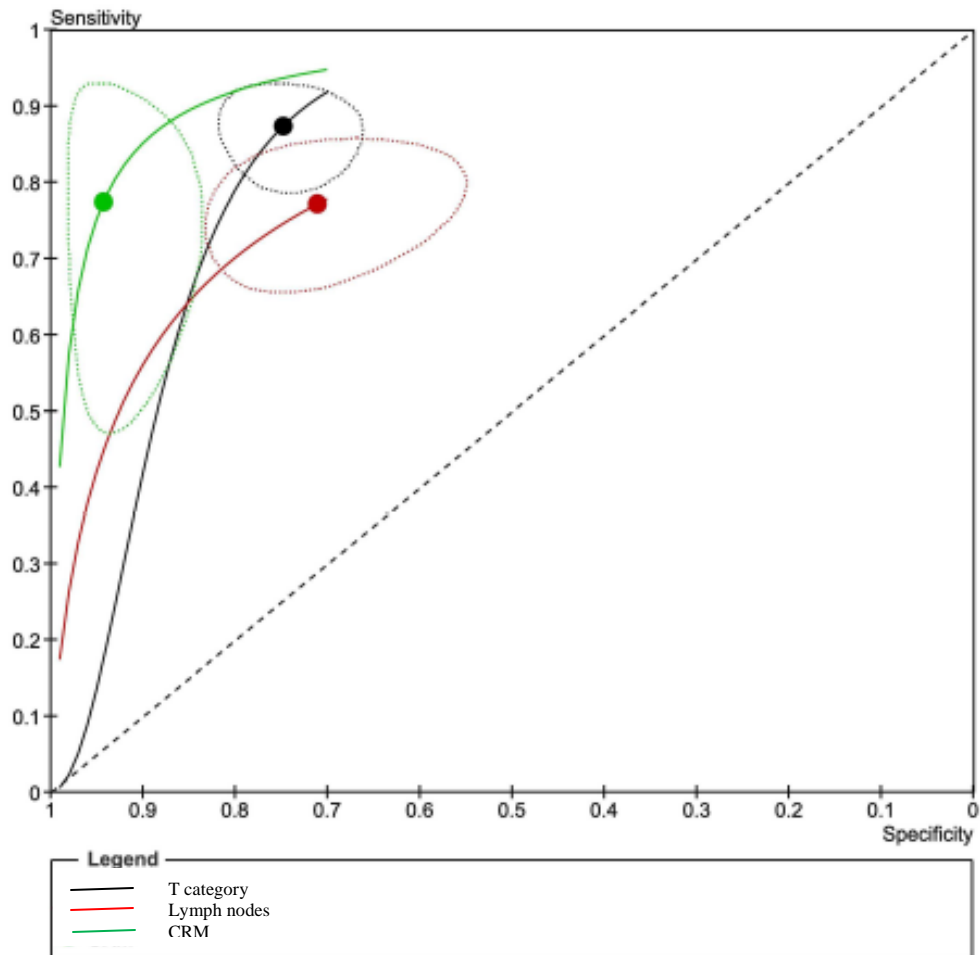
4.5 Covariate analyses

There was sufficient data to perform covariate analyses to explore the influence of approach to image assessment (independent versus consensus), MRI field strength, and adequacy of MRI protocol (adequate versus inadequate or unknown) on T-category assessment (TABLE 4.4). Studies which employed a consensus approach to image assessment produced significantly higher estimates of sensitivity (p=0.03) and DOR (p=0.02) than those in which images were reviewed independently. MRI field strength also influenced estimates, as studies using 3T machines increased the specificity of MRI compared with 1T or 1.5T machines (p=0.03). MRI protocol did not have a significant impact on MRI performance. This may have been due to the large number of studies which did not report sufficient protocol details and were therefore classified as "unknown" for this variable.

Table 4.4 Covariate analyses for T-category

Variable	Estimate			P value
Reader approach		**Independent**	**Consensus**	
	Sensitivity (%)	85 (95%CI 78-90)	94 (95%CI 85-98)	0.03
	Specificity (%)	72 (95%CI 65-79)	81 (95%CI 68-89)	0.18
	DOR	15 (95%CI 8-27)	70 (95%CI 20-237)	0.02
MRI field strength		**1T or 1.5T**	**3T**	
	Sensitivity (%)	88 (95%CI 79-93)	86 (95%CI 80-90)	0.67
	Specificity (%)	69 (95%CI 63-75)	83 (95%CI 67-92)	0.03
	DOR	16 (95%CI 8-33)	30 (95%CI 11-83)	0.32
MRI protocol		**Inadequate or unknown**	**Adequate**	
	Sensitivity (%)	88 (95%CI 76-94)	87 (95%CI 78-92)	0.90
	Specificity (%)	71 (95%CI 61-80)	76 (95%CI 68-83)	0.43
	DOR	17 (95%CI 7-45)	21 (95%CI 10-46)	0.76

Abbreviations: CI, confidence interval; CRM, circumferential resection margin; DOR, diagnostic odds ratio; T, Tesla

There were insufficient numbers to introduce other covariates into the statistical model as planned because there were too few studies in a given subgroup to allow for adequate comparison between subgroups (e.g., only four datasets defined spiculations into the perirectal fat as T2 tumour versus 21 datasets which defined spiculations as T3 tumour) or because information was not provided or unclear in a large proportion of studies. This was a particular problem with respect to covariate analyses planned for lymph node and CRM involvement since the total sample sizes for these outcomes were already relatively small (compared with T-category). However, since the widest confidence intervals were associated with sensitivity of

CRM involvement and the relevant SROC plot suggested that this may be due to three outlying studies (90,95,98) (FIGURE 4.4), a post-hoc sensitivity analysis was performed to investigate the influence of excluding these three studies on MRI sensitivity for CRM involvement. This analysis produced a higher estimate of sensitivity but had little effect on the width of the CI (93%, 95%CI 64-99).

4.6 Other findings from the systematic review

Our review identified a number of additional findings which could not be synthesized in the meta-analysis. One interesting finding was that the extramural depth of invasion (EMD) (i.e. depth of tumour invasion beyond the muscularis propria) is used in some centres as an alternative to T-category, such that patients with EMD $\geq$5mm receive preoperative CRT while those with EMD<5mm without CRM involvement proceed directly to surgery (105). The same authors showed that measurements of EMD on MRI and histopathology are equivalent within a mean difference of $\pm$0.05mm (103). However, agreement between MRI and histopathology measurements decreased for EMD>5mm (106), and a recent study (98) found that MRI determination of EMD status ($\leq$5mm versus 5mm) was no more accurate than T-category (T3 versus T1-2) assessment (88% and 91%, respectively).

With respect to CRM assessment, Beets-Tan et al retrospectively investigated the correlation between CRM measurement on phased array MRI and pathology and concluded that MRI CRM of at least 5mm is required to predict with high certainty a histologic CRM of at least 1mm (107). This study suggested that MRI prediction of CRM involvement is reliable but that accurate assessment may require the use of a wider threshold on MRI compared to pathology. However, a recent prospective study by Taylor et al showed that a cutoff of >1mm on MRI can be used to predict clear margins with a low positive histologic CRM rate (3.3%) (105).

Although short axis diameter >5mm was the criterion most commonly used to predict lymph node metastases on MRI, our review found little evidence to support this particular cut-off. In fact, one study found that 15% of lymph nodes $\leq$ 5 mm on MRI were involved with metastatic disease (108), suggesting that there is no size limit below which nodal metastasis can be ruled out. Studies that have specifically investigated optimal criteria for detecting metastatic nodes have demonstrated improved sensitivity and specificity with use of border and/or signal criteria

relative to size alone (108-110); however, such criteria can be more subjective and less reliable among different observers (97).

Chapter 5
Discussion

5.1 Interpretation of results

The results of this meta-analysis show that MRI is more specific for CRM involvement than for T-category while it is more sensitive for T-category than for CRM involvement. Previous meta-analyses have suggested that MRI may be superior for CRM assessment than for T-category, but these studies did not directly compare the two elements (52,53,64). The slight discrepancy in conclusions is likely attributable to differences in study inclusion criteria and statistical approach.

In contrast to its performance for T-category and CRM, MRI performance was more consistently poor for detection of lymph node metastases. These results are comparable to those of previous meta-analyses (52,53,63) and suggest an inherent limitation in the capacity of current MRI technology to detect lymph node metastases accurately. This problem is exacerbated by the lack of agreement on optimal criteria to assess lymph nodes. A new approach utilizing ultra-small superparamagnetic iron oxide (USPIO) to enhance nodes on MRI is under investigation and has shown promising results in rectal cancer (111-113).

While our results are largely similar to those of previous meta-analyses, the unique strength of this study is the comprehensive approach to the management of heterogeneity which has followed the meta-analysis guidelines set by the Cochrane Collaboration. This has been accomplished in three main ways. First, focused inclusion criteria have been applied to minimize differences among study populations and technology. Second, a rigorous statistical approach has been utilized to incorporate heterogeneity into estimates and reflect its extent in the range of possible values (95% confidence intervals), with covariate analyses performed when possible to explore the role of different variables. Third, a formal assessment of methodological quality using a validated tool has been performed to describe potential sources of heterogeneity. Thus, the diagnostic accuracy estimates produced in this analysis represent the highest quality evidence available at the present time for MRI assessment of T-category, lymph node metastases, and CRM involvement in rectal cancer.

A new finding of this study which has not been identified in previous meta-analyses is that T-category assessment is improved with the use of higher field strength (3T) MRI machines and review of images by consensus of two or more radiologists. This finding is important because both variables are potentially modifiable and can be directly translated into clinical practice. Image assessment by consensus produced the highest sensitivity estimate in covariate analyses and is of particular clinical value because it represents a simple intervention that may be implemented using existing infrastructure and personnel to improve the diagnostic utility of available MRI technology.

The improvement of MRI diagnostic accuracy through consensus reading is consistent with evidence that inter-observer agreement between radiologists is highly variable for rectal cancer staging. One of the studies included in our meta-analysis measured inter-observer agreement among four gastrointestinal radiologists for detecting T3 or greater versus T2 or lesser tumours (89). Agreement ranged from moderate to excellent (kappa=0.53-0.82) for the different pairs of readers. Individual reader experience was not described and it was unclear to what extent it may have influenced results. In two other studies which compared two radiologists with varying expertise in gastrointestinal imaging, agreement on T-category was moderate both for distinguishing T3 or greater versus T2 or lesser tumours (linear weighted kappa=0.47, 95%CI 0.31-0.64) (96) as well as for identification of individual T1-T4 categories (linear weighted kappa=0.53, 95%CI 0.38-0.69) (107). Such variability exists not only for T-category but also for CRM and lymph node assessment (89,96), and this suggests that a consensus-based approach to image review may improve the accuracy of MRI for these elements as well (though this could not be directly investigated in our dataset due to insufficient numbers). Since the results published in the diagnostic accuracy literature are typically based on image assessment by experienced GI radiologists, true variability in general clinical practice is likely even higher, making the role of multiple readers even more valuable.

Evidence of the value of consensus assessment is limited in the radiology literature; however, the problem of inter-observer variation has been extensively investigated in other diagnostic disciplines that rely on clinician interpretation of test or specimen results. In the field of pathology, for example, variability has been demonstrated in the assessment of histopathology specimens for the diagnosis or staging of numerous disease processes, both benign and malignant

(114-118). In some instances, such discrepancies had the potential to alter clinical decision-making significantly. As part of its guidelines for quality assurance, the College of American Pathologists (CAP) has addressed this issue through a number of recommendations that include the establishment of daily intradepartmental conferences wherein pathologic assessment for difficult cases may be reviewed and a final diagnosis achieved by consensus (119). In the same vein, the American Society of Clinical Pathologists (ASCP) recommends the institution of formal mechanisms for second opinion in diagnostic surgical pathology, with special attention given to cases identified as "problem-prone" based on personal, local, or literature experience, as well as highly critical or significant cases such as new malignant neoplasms or cases in which irreversible surgery will be performed (120). In Ontario, second opinion consultations were recently proposed as part of provincial recommendations to improve the quality of pathology assessment (121,122). Pathology intradepartmental consensus conferences have been successfully implemented at major centres in North America and worldwide (123).

5.2 Limitations

While this study has employed a thorough approach to the management of heterogeneity, there remain a number of sources of variability which could not be completely taken into account in our analysis. These included clinical definitions of staging elements, approach to preoperative treatment decisions, MRI protocols, radiologist experience, and details of study methodology. We had planned to explore the impact of these individual variables on heterogeneity through covariate analyses; however, this could not be completed due to insufficient numbers and incomplete reporting of study details. For example, statistical heterogeneity was greatest in CRM sensitivity estimates and did not appear to diminish when the three studies with low sensitivity (90,95,98) were excluded in a sensitivity analysis, suggesting that the variability could not be explained by these outlying points alone. A threshold effect related to different definitions of CRM involvement is likely an important contributor to this heterogeneity, as wider cutoffs produced higher sensitivity and lower specificity estimates, while the reverse was true for narrower cutoffs. The lack of complete reporting of study details is well recognized as an important limitation in the evaluation of diagnostic accuracy literature (124). Our findings further emphasize the need for academic publications to adopt and enforce uniform and complete

reporting standards, such as those recommended by the Cochrane Collaboration in the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) statement (125).

A second limitation of this analysis is the exclusion of patients who underwent preoperative CRT. Since preoperative CRT was allocated on the basis of MRI findings (using T-category and lymph nodes in some cases, CRM in others), excluding those patients could result in apparently lower sensitivity and higher specificity, as analysis could identify cases that were understaged (i.e. those who should have received CRT but did not) but could fail to identify most of those that were overstaged (i.e. those who should not have received CRT but did). Including these patients would be equally problematic, however, as CRT administered following MRI might downstage the tumour so that final histopathology no longer represents an appropriate reference standard for the original MR image. This problem applies to all meta-analyses of this topic. To determine the extent to which sensitivity and specificity were influenced by the strategy for allocation of CRT, it would be necessary to examine individual studies for their criteria for administering CRT, the proportion of patients who underwent this treatment, and the results of histopathology in those patients. Such analysis was not possible in our review, as much of this information was not available.

A third limitation is that neither the bivariate nor HSROC models currently incorporate methods to account for the inclusion of multiple datasets from a single study. In this meta-analysis, studies which reported multiple datasets (e.g. from multiple independent observers) were treated as separate datasets, but it is reasonable to assume that there is some degree of correlation between them due to similarities in population, technique, or interpretation. Accounting for correlation in the data would help to produce more accurate estimates of variance and thus reduce the degree of uncertainty around mean sensitivity and specificity estimates (i.e., narrower confidence intervals). This problem must be addressed in future formulations of these models.

Chapter 6
Summary and future directions

MRI is very specific for assessment of CRM involvement but is more sensitive for T-category. CRM and T-category are likely complementary indicators of the need for preoperative CRT and both should be taken into account. Using MRI with higher field strength and reviewing images in consensus with a second radiologist are strategies that can increase the sensitivity and specificity of MRI for T-category assessment and should be considered as interventions to improve the diagnostic value of available MRI technology. In contrast, lymph node assessment is poor on MRI and approaches to improve it should be further investigated.

Heterogeneity remains a significant problem for meta-analyses of diagnostic accuracy studies and arises from differences in populations, methodology, and test parameters including definitions of positive cases. In contrast to previous meta-analyses on this topic, heterogeneity was managed in this meta-analysis in a number of ways including 1) restrictive eligibility criteria, 2) use of a rigorous statistical approach which took into account heterogeneity both within (i.e., sampling variability) and between studies, and 3) formal quality assessment using a validated tool. Definitions of essential staging elements and MRI protocols should be standardized to reduce heterogeneity in future studies.

The optimal approach to preoperative staging and neoadjuvant treatment allocation in rectal cancer remains a matter of controversy. A recent review criticized the practice of using MRI to predict CRM due to "methodological inadequacies" in diagnostic accuracy studies published to date on this subject (126). These included discrepancies in definitions of CRM involvement between MRI and pathology and lack of blinding of surgeons and pathologists to MRI findings. As a next step in addressing this problem, our analysis can provide the foundation to support a randomized controlled trial of MRI-based preoperative staging and treatment allocation in rectal cancer that compares traditional TNM criteria to a combination of T-category and CRM. Such a study would allow direct comparison of survival and local recurrence outcomes between the two approaches and would provide a definitive answer to this critical question.

References

1.   Statistics Canada, Canadian Cancer Registry (CCR) Database, July 2010. Available from http://www.statcan.gc.ca/pub/82-231-x/2009001/t029-eng.htm

2.   Kapiteijn E, Marijnen CAM, Nagtegaal ID, Putter H, Steup WH, Wiggers T, et al. Preoperative radiotherapy combined with mesorectal excision for resectable rectal cancer. N Engl J Med. 2001 August 30;345(9):638-646.

3.   MacFarlane JK, Ryall RDH, Heald RJ. Mesorectal excision for rectal cancer. Lancet. 1993 February 20;341(8843):457-460.

4.   Yun H, Lee L, Park J, Cho Y, Cho Y, Lee W, et al. Local recurrence after curative resection in patients with colon and rectal cancers. Int J Colorectal Dis. 2008 November;23(11):1081-1087.

5.   Manfredi S, Benhamiche AM, Meny B, Cheynel N, Rat P, Faivre J. Population-based study of factors influencing occurrence and prognosis of local recurrence after surgery for rectal cancer. Br J Surg. 2001 September;88(9):1221-1227.

6.   Palmer G, Martling A, Cedermark B. A population-based study on the management and outcome in patients with locally recurrent rectal cancer. Ann Surg Oncol. 2007 February;14(2):447-454.

7.   Miner TJ, Jaques DP, Paty PB, Guillem JG, Wong WD. Symptom control in patients with locally recurrent rectal cancer. Ann Surg Oncol. 2003 January-February;10(1):72-79.

8.   Esnaola NF, Cantor SB, Johnson ML, Mirza AN, Miller AR, Curley SA, et al. Pain and quality of life after treatment in patients with locally recurrent rectal cancer. J Clin Oncol. 2002 November 1; 20(21):4361-4367.

9.   Jörgren F, Johansson R, Damber KL, Lindmark G. Risk factors of rectal cancer local recurrence: population-based survey and validation of the Swedish rectal cancer registry. Colorectal Dis. 2010 October;12(10):977-986.

10.   Pilipshen SJ, Heilweil M, Quan SH, Sternberg SS, Enker WE. Patterns of pelvic recurrence following definitive resections of rectal cancer. Cancer. 1984 March 15;53(6):1354-1362.

11.   Zaheer S, Pemberton JH, Farouk R, Dozois RR, Wolff BG, Ilstrup D. Surgical treatment of adenocarcinoma of the rectum. Ann Surg. 1998 June;227(6):800–811.

12.   Jatzko GR, Jagoditsch M, Lisborg PH, Denk H, Klimpfinger M, Stettner HM. Long-term results of radical surgery for rectal cancer: multivariate analysis of prognostic factors influencing survival and local recurrence. Eur J Surg Oncol. 1999 June; 25(3):284-291.

13.   Bokey EL, Ojerskog B, Chapuis PH, Dent OF, Newland RC, Sinclair G. Local recurrence after curative excision of the rectum for cancer without adjuvant therapy: role of total anatomical dissection. Br J Surg. 1999 September 1;86(9):1164-1170.

14.   Hohenberger W, Schick CH, Gohl J. Mesorectal lymph node dissection: is it beneficial? Langenbecks Arch Surg 1998 December;383(6):402-408.

15.   Wibe A, Rendedal PR, Svensson E, et al. Prognostic significance of the circumferential resection margin following total mesorectal excision for rectal cancer. Br J Surg 2002 March;89(3):327–334.

16.   Birbeck KF, Macklin CP, Tiffin NJ, Parsons W, Dixon MF, Mapstone NP, et al. Rates of circumferential resection margin involvement vary between surgeons and predict outcomes in rectal cancer surgery. Ann Surg. 2002 April;235(4):449–457.

17.   Adam IJ, Martin IG, Finan PJ, Johnston D, Mohamdee MO, Scott N, et al. Role of circumferential resection margin involvement in the local recurrence of rectal cancer. Lancet. 1994 September 10; 344(8924):707-711.

18.   Quirke P, Durdey P, Dixon MF, Williams NS. Local recurrence of rectal adenocarcinoma due to inadequate surgical resection: histopathological study of lateral tumour spread and surgical excision. Lancet. 1986 November 1;2(8514):996-999.

19.   Cawthorn SJ, Parums DV, Gibbs NM, A'Hern RP, Caffarey SM, Broughton CI, et al. Extent of mesorectal spread and involvement of lateral resection margin as prognostic factors after surgery for rectal cancer. Lancet. 1990 May 5; 335(8697):1055–1059.

20.   Hall NR, Finan PJ, Al-Jaberi T, Tsang CS, Brown SR, Dixon MF, et al. Circumferential margin involvement after mesorectal excision of rectal cancer with curative intent: predictor of survival but not local recurrence? Dis Colon Rectum. 1998 August;41(8):979-983.

21.   Bleday R, Brindzei N. Surgical treatment of rectal cancer. In: Beck DE, Roberts PL, Saclarides TJ, Senagore AJ, Stamos MJ, Wexner SD, eds. The ASCRS Textbook of Colon and Rectal Surgery, 2nd edition. New York: Springer;2011:743-760.

22.   Hurst PA, Prout WG, Kelly JM, Bannister JJ, Walker RT. Local recurrence after low anterior resection using the staple gun. Br J Surg. 1982 May;69(5):275-276.

23.   Lasson ALL, Ekelund GR, Lindstrom CG. Recurrence risk after stapled anastomosis for rectal carcinoma. Acta Chir Scand. 1984;150(1):85-89.

24.   Heald RJ, Husband EM, Ryall RD. The mesorectum in rectal cancer surgery: the clue to pelvic recurrence? Br J Surg. 1982 October;69(10):613-616.

25.   Heald RJ. Rectal cancer: the surgical options. Eur J Cancer. 1995 July-August;31A(7-8):1189-1192.

26.   Enker WE. Potency, cure, and local control in the operative treatment of rectal cancer. Arch Surg. 1992 December;127(12):1396-1401.

27.   Havenga K, Enker WE, Norstein J, Moriya Y, Heald RJ, van Houwelingen HC, et al. Improved survival and local control after total mesorectal excision or D3 lymphadenectomy in the treatment of primary rectal cancer: An international analysis of 1411 patients. Eur J Surg Oncol. 1999 August;25(4):368-374.

28.   Quirke P, Dixon MF. The prediction of local recurrence in rectal adenocarcinoma by histopathological examination. Int J Colorectal Dis. 1988 June;3(2):127-131.

29.   Maslekar S, Sharma A, MacDonald A, Gunn J, Monson JRT, Hartley JE. Mesorectal grades predict recurrences after curative resection for rectal cancer. Dis Colon Rectum. 2007 February;50(2):168-175.

30.   Quirke P, Steele R, Monson J, Grieve R, Khanna S, Couture J, et al. Effect of the plane of surgery achieved on local recurrence in patients with operable rectal cancer: a prospective study using data from the MRC CR07 and NCIC-CTG CO16 randomised clinical trial. Lancet. 2009 March;373(9666):821-828.

31.   Swedish Rectal Cancer Trial. Improved survival with preoperative radiotherapy in resectable rectal cancer. N Engl J Med. 1997 April 3;336(14):980-987.

32.   Peeters KC, Marijnen CA, Nagtegaal ID, Kranenbarg EK, Putter H, Wiggers T, et al. The TME trial after a median follow-up of 6 years: increased local control but no survival

benefit in irradiated patients with resectable rectal carcinoma. Ann Surg. 2007 November;246(5):693-701.

33.   Bujko K, Nowacki MP, Nasierowska-Guttmejer A, Michalski W, Bebenek M, Kryj M. Long-term results of a randomized trial comparing preoperative short-course radiotherapy with preoperative conventionally fractionated chemoradiation for rectal cancer. Br J Surg. 2006 October;93(10):1215-1223.

34.   Gérard JP, Conroy T, Bonnetain F, Bouché O, Chapet O, Closon-Dejardin MT, et al. Preoperative radiotherapy with or without concurrent fluorouracil and leucovorin in T3-4 rectal cancers: results of FFCD 9203. J Clin Oncol. 2006 October 1;24(28):4620-4625.

35.   Bosset JF, Collette L, Calais G, Mineur L, Maingon P, Radosevic-Jelic L, et al. Chemotherapy with preoperative radiotherapy in rectal cancer. N Engl J Med. 2006 September 14;355(11):1114-1123.

36.   Sauer R, Becker H, Hohenberger W, Rödel C, Wittekind C, Fietkau R, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med. 2004 October 21;351(17):1731-1740.

37.   Frykholm GJ, Glimelius B, Påhlman L. Preoperative or postoperative irradiation in adenocarcinoma of the rectum: final treatment results of a randomized trial and an evaluation of late secondary effects. Dis Colon Rectum. 1993 June;36(6):564-572.

38.   Colorectal Cancer Collaborative Group. Adjuvant radiotherapy for rectal cancer: a systematic overview of 8,507 patients from 22 randomised trials. Lancet 2001 October 20;358(9290):1291-1304.

39.   National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology available at http://www.nccn.org/professionals/physician_gls/f_guidelines.asp (Accessed September 10, 2011).

40.   Wong RKS, Berry S, Spithoff K, Simunovic M, Chan K, Agboola O, et al. Postoperative therapy for stage II or III rectal cancer: an updated practice guideline. Clin Oncol. 2010 May;22(4):265-271.

41.   Marijnen CA, Kapiteijn E, van de Velde CJ, Martijn H, Steup WH, Wiggers T, et al. Acute side effects and complications after short-term preoperative radiotherapy combined with total mesorectal excision in primary rectal cancer: report of a multicenter randomized trial. J Clin Oncol. 2002 February 1;20(3):817-825.

42.   Marijnen CA, van de Velde CJ, Putter H, van den Brink M, Maas CP, Martijn H, et al. Impact of short-term preoperative radiotherapy on health-related quality of life and sexual functioning in primary rectal cancer: report of a multicenter randomized trial. J Clin Oncol. 2005 March 20;23(9):1847-1858.

43.   Peeters KC, van de Velde CJ, Leer JW, Martijn H, Junggeburt JM, Kranenbarg EK, et al. Late side effects of short-course preoperative radiotherapy combined with total mesorectal excision for rectal cancer: increased bowel dysfunction in irradiated patients - a Dutch colorectal cancer group study. J Clin Oncol. 2005 September 1;23(25):6199-6206.

44.   Pollack J, Holm T, Cedermark B, Altman D, Holmström B, Glimelius B, et al. Late adverse effects of short-course preoperative radiotherapy in rectal cancer. Br J Surg. 2006 December;93(12):1519-1525.

45.   Pietrzak L, Bujko K, Nowacki MP, Kepka L, Oledzki J, Rutkowski A, et al. Quality of life, anorectal and sexual functions after preoperative radiotherapy for rectal cancer: report of a randomised trial. Radiother Oncol. 2007 September;84(3):217-225.

46.   Merkel S, Mansmann U, Siassi M, Papadopoulos T, Hohenberger W, Hermanek P. The prognostic inhomogeneity in pT3 rectal carcinomas. Int J Colorectal Dis. 2001 September;16(5):298-304.

47.   Carne PWG, Nelson H. Are we overtreating rectal cancer: time for another trial? Ann Surg Oncol. 2004 February;11(2):124-126.

48.   Pahlman L. Indications for long-term radiotherapy. Recent Results Cancer Res. 2005;165:212-220.

49.   MacKay G, Downey M, Molloy RG, O'Dwyer RJ. Is preoperative radiotherapy necessary in T1–T3 rectal cancer with TME? Colorectal Dis. 2006 January;8(1):34-36.

50.   Krook JE, Moertel CG, Gunderson LL, Wieand HS, Collins RT, Beart RW et al. Effective surgical adjuvant therapy for high-risk rectal carcinoma. N Engl J Med. 1991 March 14;324(11):709-715.

51.   Brown G, Davies S, Williams GT, Bourne MW, Newcombe RG, Radcliffe AG, et al. Effectiveness of preoperative staging in rectal cancer: digital rectal examination, endoluminal ultrasound or magnetic resonance imaging? Br J Cancer. 2004 July 5;91(1):23-29.

52.   Kwok H, Bissett IP, Hill GL. Preoperative staging of rectal cancer. Int J Colorectal Dis. 2000 February;15(1):9-20.

53.   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 September;232(3):773-783.

54.   Rosen MP, Bree RL, Foley WD, Gay SB, Grant TH, Heiken JP, et al. Expert Panel on Gastrointestinal Imaging. ACR Appropriateness Criteria pretreatment staging of colorectal cancer. [online publication]. Reston (VA): American College of Radiology (ACR); 2008.

55.   Augestad KM, Lindsetmo RO, Stulberg J, Reynolds H, Senagore A, Champagne B, et al. International preoperative rectal cancer management: staging, neoadjuvant treatment, and impact of multidisciplinary teams. World J Surg. 2010 November;34(11):2689-2700.

56.   Chern H, Wong WD. Clinical staging: endoscopic techniques. In: Czito B, Willett CG, eds. International Perspectives on Multimodality Management. New York: Humana Press;2010:1-20.

57.   Beets-Tan RG. MRI in rectal cancer: the T stage and circumferential resection margin. Colorect Dis. 2003 September;5(5):392-395.

58.   Brown G, Daniels IR. Preoperative staging of rectal cancer: the MERCURY research project. Recent Results Cancer Res. 2005;165:58-74.

59.   Brown G, Kirkham A, Williams GT, Bourne M, Radcliffe AG, Sayman J, et al. High-resolution MRI of the anatomy important in total mesorectal excision of the rectum. AJR Am J Roentgenol. 2004 February;182(2):431-439.

60.   Martling A, Holm T, Bremmer S, Lindholm J, Cedermark B, Blomqvist L. Prognostic value of preoperative magnetic resonance imaging of the pelvis in rectal cancer. Br J Surg. 2003 November;90(11):1422-1428.

61.   Wieder HA, Rosenberg R, Lordick F, Geinitz H, Beer A, Becker K, et al. Rectal cancer: MR Imaging before neoadjuvant chemotherapy and radiation therapy for prediction of tumor-free circumferential resection margins and long-term survival. Radiology. 2007 June;243(3):744-751.

62.   Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane

Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org.

63.   Lahaye MJ, Engelen SME, Nelemans PJ, Beets GL, van de Velde CJ, van Engelshoven JM, et al. Imaging for predicting the risk factors – the circumferential resection margin and nodal disease – of local recurrence in rectal cancer: a meta-analysis. Semin Ultrasound CT MRI. 2005 August;26(4):259-268.

64.   Purkayastha S, Tekkis PP, Athanasiou T, Tilney HS, Darzi AW, Heriot AG. Diagnostic precision of magnetic resonance imaging for preoperative prediction of the circumferential margin involvement in patients with rectal cancer. Colorect Dis. 2007 June;9(5):402-411.

65.   Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2009. Available from: http://srdta.cochrane.org/.

66.   Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003 November 10;3:25.

67.   Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press, 1995.

68.   Jadad AR, Moore A, Caroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996 February;17(1):1-12.

69.   Whiting P, Westwood M, Rutjes AWS, Reitsma J, Bossuyt P, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol. 2006 March 6;6:9.

70.   Henschke N, Maher C, Refshauge K. A systematic review identified five "red flags" to screen for vertebral fracture in patients with low back pain. J Clin Epi. 2008 February;61(2):110-118.

71.   Stengel D, Bauwens K, Rademacher G, Mutze S, Ekkernkamp A. Association between compliance with methodological standards of diagnostic research and reported test

accuracy: meta-analysis of focused assessment of US for trauma. Radiology. 2005 July;236(1):102-111.

72. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. 2003 November;56(11):1129-1135.

73. Brenner H, Gefeller O. Variations of sensitivity, specificity, likelihood ratios, and predictive values with disease prevalence. Stat Med. 1997 May;16(9):981-991.

74. Li J, Fine JP. Assessing the dependence of sensitivity and specificity on prevalence in meta-analysis. Biostatistics 2011 October;12(4):710-722.

75. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

76. Harbord RM, Whiting P, Sterne JAC, Egger M, Deeks JJ, Shang A, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. J Clin Epi. 2008 November;61(11):1095-1103.

77. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making. 1993 October-December;13(4):313-321.

78. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993 July 30;12(14):1293-1316.

79. Gatsonis C, Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. AJR Am J Roentgenol. 2006 August;187(2):271-281.

80. Arends LR, Hamza TH, Houwelingen JC, Heijenbrok-Kal MH, Hunink MG, Stijnen T. Bivariate random effects meta-analysis of ROC curves. Med Decis Mak. 2008 September-October;28(5):621-638.

81. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005 October;58(10):982-990.

82. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001 October 15;20(19):2865-2884.

83.   Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics. 2007 April;8:239-251.

84.   Branagan G, Chave H, Fuller C, McGee S, Finnis D. Can magnetic resonance imaging predict circumferential margins and TNM stage in rectal cancer? Dis Colon Rectum. 2004 August;47(8):1317-1322.

85.   Burton S, Brown G, Daniels I, Norman A, Swift I, Abulafi M, et al. MRI identified prognostic features of tumors in distal sigmoid, rectosigmoid, and upper rectum: treatment with radiotherapy and chemotherapy. Int J Radiat Oncol Biol Phys. 2006 June 1;65(2):445-451.

86.   Ferri M, Laghi A, Mingazzini P, Iafrate F, Meli L, Ricci F, et al. Pre-operative assessment of extramural invasion and sphincteral involvement in rectal cancer by magnetic resonance imaging with phased-array coil. Colorect Dis. 2005 July;7(4):387-393.

87.   Gagliardi G, Bayar S, Smith R, Salem RR. Preoperative staging of rectal cancer using magnetic resonance imaging with external phase-arrayed coils. Arch Surg. 2002 April;137(4):447-451.

88.   Halefoglu AM, Yildirim S, Avlanmis O, Sakiz D, Baykan A. Endorectal ultrasonography versus phased-array magnetic resonance imaging for preoperative staging of rectal cancer. World J Gastroenterol. 2008 June 14;14(22):3504-3510.

89.   Kim MJ, Lim JS, Oh YT, Kim JH, Chung JJ, Joo SH, et al. Preoperative MRI of rectal cancer with and without rectal water filling: an intraindividual comparison. AJR Am J Roentgenol. 2004 June;182(6):1469-1476.

90.   MERCURY Study Group. Diagnostic accuracy of preoperative magnetic resonance imaging in predicting curative resection of rectal cancer: prospective observational study. BMJ. 2006 October 14;333(7572):779-784.

91.   Oberholzer K, Junginger T, Kreitner KF, Krummenauer F, Simiantonaki N, Trouet S, et al. Local staging of rectal carcinoma and assessment of the circumferential resection margin with high-resolution MRI using an integrated parallel acquisition technique. J Magn Reson Imaging. 2005 July;22(1):101-108.

92.   Piippo U, Paakko E, Makinen M, Makela J. Local staging of rectal cancer using the black lumen magnetic resonance imaging technique. Scand J Surg. 2008;97(3):237-242.

93. Rao SX, Zeng MS, Xu JM, Qin XY, Chen CZ, Li RC, et al. Assessment of T staging and mesorectal fascia status using high-resolution MRI in rectal cancer with rectal distention. World J Gastroenterol. 2007 August 14;13(30):4141-4146.

94. Strassburg J, Lewin A, Ludwig K, Killian L, Linke J, Loy V, et al. Optimised surgery (so-called TME surgery) and high-resolution MRI in the planning of treatment of rectal carcinoma. Langenbecks Arch Surg. 2007 March;392(2):179-188.

95. Taylor A, Slater A, Mapstone N, Taylor S, Halligan S. Staging rectal cancer: MRI compared to MDCT. Abdom Imaging. 2007 May-June;32(3):323-327.

96. Vliegen RFA, Beets GL, von Meyenfeldt MF, Kessels AG, Lemaire EE, van Engelshoven JM, et al. Rectal cancer: MR imaging in local staging - is gadolinium-based contrast material helpful? Radiology. 2005 January;234(1):179-188.

97. Akasu T, Iinuma G, Takawa M, Yamamoto S, Muramatsu Y, Moriyama N. Accuracy of high-resolution magnetic resonance imaging in preoperative staging of rectal cancer. Ann Surg Oncol. 2009 October;16(10):2787-2794.

98. Kim YW, Cha SW, Pyo J, Kim NK, Min BS, Kim MJ, et al. Factors related to preoperative assessment of the circumferential resection margin and the extent of mesorectal invasion by magnetic resonance imaging in rectal cancer: a prospective comparison study. World J Surg. 2009 September;33(9):1952-1960.

99. Kim CK, Kim SH, Chun HK, Lee WY, Yun SH, Song SY, et al. Preoperative staging of rectal cancer: accuracy of 3-Tesla magnetic resonance imaging. Eur Radiol 2006 May;16(5):972-980.

100. Kam MH, Wong DC, Stevenson ARL, Lai K, Phillips GE. Comparison of magnetic resonance imaging–fluorodeoxyglucose positron emission tomography fusion with pathological staging in rectal cancer. Br J Surg. 2010 February;97(7):266-268.

101. Kim H, Lim JS, Choi JY, Park J, Chung YE, Kim MJ, et al. Rectal Cancer: comparison of accuracy of local-regional staging with two- and three-dimensional preoperative 3-T MR imaging. Radiology. 2010 February;254(2):485-492.

102. Kim SH, Lee JM, Lee MW, Kim GH, Han JK, Choi BI. Diagnostic accuracy of 3.0-Tesla rectal magnetic resonance imaging in preoperative local staging of primary rectal cancer. Invest Radiol. 2008 August;43(8):587-593.

103. Futterer JJ, Yakar D, Strijk SP, Barentsz JO. Preoperative 3 T MR imaging of rectal cancer: local staging accuracy using a two-dimensional and three-dimensional T2-weighted turbo spin echo sequence. Eur J Radiol. 2008 January;65(1):66-71.

104. MERCURY Study Group. Extramural depth of tumor invasion at thin-section MR in patients with rectal cancer: results of the MERCURY study. Radiology. 2007 April;243(1):132-139.

105. Taylor FGM, Quirke P, Heald RJ, Moran B, Blomqvist L, Swift I, et al. Preoperative high-resolution magnetic resonance imaging can identify good prognosis stage I, II, and III rectal cancer best managed by surgery alone: a prospective, multicenter, European study. Ann Surg. 2011 April;253(4):711-719.

106. Park SH. Degree of error of thin-section MR in measuring extramural depth of tumor invasion in patients with rectal cancer. Radiology. 2008 February;246(2):647-648.

107. Beets-Tan RGH, Beets GL, Vliegen RFA, Kessels AG, Van Boven H, De Bruine A, et al. Accuracy of magnetic resonance imaging in prediction of tumor-free resection margin in rectal cancer surgery. Lancet. 2001 February;357(9255):497-504.

108. Brown G, Richards CJ, Bourne MW, Newcombe RG, Radcliffe AG, Dallimore NS, et al. Morphologic predictors of lymph node status in rectal cancer with use of high-spatial-resolution MR imaging with histopathologic comparison. Radiology. 2003 May;227(2):371-377.

109. Kim JH, Beets GL, Kim MJ, Kessels AG, Beets-Tan RG. High-resolution MR imaging for nodal staging in rectal cancer: are there any criteria in addition to the size? Eur J Radiol. 2004 October;52(1):78-83.

110. Matsuoka H, Nakamura A, Sugiyama M, Hachiya J, Atomi Y, Masaki T. MRI diagnosis of mesorectal lymph node metastasis in patients with rectal carcinoma: what is the optimal criterion? Anticancer Res. 2004 November-December;24(6):4097-4101.

111. Koh DM, Brown G, Temple L, Raja A, Toomey P, Bett N, et al. Rectal Cancer: mesorectal lymph nodes at MR imaging with USPIO versus histopathologic findings - initial observations. Radiology. 2004 April;231(1):91-99.

112. Lahaye MJ, Engelen SM, Kessels AG, de Bruine AP, von Meyenfeldt MF, van Engelshoven JM, et al. USPIO-enhanced MR imaging for nodal staging in patients with primary rectal cancer: predictive criteria. Radiology. 2008 March;246(3):804-811.

113. Engelen SM, Beets-Tan RG, Lahaye MJ, Kessels AG, Beets GL. Location of involved mesorectal and extramesorectal lymph nodes in patients with primary rectal cancer: preoperative assessment with MR imaging. Eur J Surg Oncol. 2008 July;34(7):776-781.

114. Kronz JD, Westra WH, Epstein JI. Mandatory second opinion surgical pathology at a large referral hospital. Cancer. 1999 December 1;86(11):2426-2435.

115. Abt AB, Abt LG, Olt GJ. The effect of interinstitution anatomic pathology consultation on patient care. Arch Pathol Lab Med. 1995 June;119(6):514-517.

116. Epstein JI, Walsh PC, Sanfilippo F. Clinical and cost impact of second opinion pathology review of prostate biopsies prior to radical prostatectomy. Am J Surg Pathol. 1996 July;20(7):851-857.

117. Selman AE, Niemann TH, Fowler JM, Copeland LJ. Quality assurance of second opinion in gynecologic oncology. Obstet Gynecol. 1999 August;94(2):302-306.

118. Harris M, Hartley AL, Blair V, Birch JM, Banerjee SS, Freemont AJ, et al. Sarcomas in north west England, I: histopathological peer review. Br J Cancer. 1991 August;64(2):315-320.

119. Nakhleh RE, Fitzgibbon PL. Quality improvement manual in anatomic pathology. Second edition. Northfield, IL: College of American Pathologists;2002.

120. Tomaszewski JE, Bear HD, Connally JA, Epstein JI, Feldman M, Foucar K, et al. Consensus conference on second opinions in diagnostic anatomic pathology. Who, what, and when. Am J Clin Pathol. 2000 September;114(3):329-35.

121. McLellan B, McLeod R, Srigley J. Report of the investigators of surgical and pathology issues at three Essex County hospitals: Hôtel-Dieu Grace Hospital, Leamington District Memorial Hospital and Windsor Regional Hospital. July 2010. [online publication].

122. McLachlin CM, Srigley JR, Mullen JBM, Chorneyko KA, Tron VA, Joshi SB, et al. Standards2Quality - Guidelines for quality management in surgical pathology professional practices: a proposal for laboratory physicians in Ontario. 2011. [online publication].

123. Khurshid A, Ahmad Z, Qureshi A. In-house daily consensus conference: An important quality control/quality assurance activity - experience at a major referral center. Indian J Pathol Microbiol. 2009 July-September;52(3):325-327.

124. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. JAMA. 1995 August;274(8):645-651.

125. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. Radiology. 2003 January;226(1):24-28.

126. Dent OF, Chapuis PH, Haboubi N, Bokey L. Magnetic resonance imaging cannot predict histological tumour involvement of a circumferential surgical margin in rectal cancer. Colorectal Dis. 2011 September;13(9):974-981.

Appendices

Table A.1 American Joint Committee on Cancer (AJCC) TNM staging classification for rectal cancer

Table A.1a Criteria for T, N, and M categories

Primary Tumour (T)	Tx	Primary tumour cannot be assessed.
	T0	No evidence of primary tumour.
	Tis	Tis Carcinoma *in situ*: intraepithelial or invasion of lamina propria.
	T1	Tumour invades submucosa.
	T2	Tumour invades muscularis propria.
	T3	Tumour invades through the muscularis propria into pericolorectal tissues.
	T4a	Tumour penetrates to the surface of the visceral peritoneum.
	T4b	Tumour directly invades or is adherent to other organs or structures.
Regional Lymph Nodes (N)	Nx	Regional lymph nodes cannot be assessed.
	N0	No regional lymph node metastasis.
	N1	Metastases in 1-3 regional lymph node.
	N2	Metastases in $\geq$4 regional lymph nodes.
Distant Metastasis (M)	M0	No distant metastasis.
	M1	Distant metastasis.

Table A.1b Stage prognostic groups

Stage	**T**	**N**	**M**
0	Tis	N0	M0
I	T1-T2	N0	M0
II	T3-T4	N0	M0
III	Any T	N1-N2	M0
IV	Any T	Any N	M1

(Source: Colon and rectum. In: Edge SB, Byrd DR, Compton CC, et al., eds.: AJCC Cancer Staging Manual. 7th ed. New York, NY: Springer, 2010, pp 143-64.)

Table A.2 Summary of randomized controlled trials in preoperative radiotherapy or chemoradiotherapy for rectal cancer

Trial	N	Description	Preoperative Staging Modality	Local Recurrence (%)		Overall Survival (%)	
				Pre-op RT	No RT	Pre-op RT	No RT
Swedish (NEJM 1997)	1168	Stage I-III Pre-op short course RT vs No Pre-op RT 5 year follow-up	Not described	11*	27*	58*	48*
Dutch (NEJM 2001)	1861	Stage I-III Pre-op short course RT vs No Pre-op RT 2 year follow-up	Not described	2*	8*	82	82
Dutch long term follow-up (Ann Surg 2005)	1861	Stage I-III Pre-op short course RT vs No Pre-op RT 6 year follow-up	Not described	5.6*	10.9*	64.2	63.5
				Pre-op CRT	Post-op CRT	Pre-op CRT	Post-op CRT
Sauer (NEJM 2004)	823	Stage II and III Pre-op CRT vs Post-op CRT 5 year follow-up	ERUS and CT	6*	13*	76	74
				Pre-op RT	Pre-op CRT	Pre-op RT	Pre-op CRT
Bujko (BJS, 2006)	316	T3 or resectable T4 with no distant metastases Pre-op short course RT vs Pre-op CRT 4 year follow-up	DRE, CT, ERUS and/or MRI	9.0	14.2	67	66
Gerard (JCO 2006)	762	T3 or resectable T4 with no distant metastases Pre-op long course RT vs Pre-op CRT 5 year follow-up	ERUS and CT	16.5*	8.1*	67.9	67.4
Bosset (NEJM 2006)	1011	T3 or resectable T4 with no distant metastases Four treatment arms: 1) Pre-op long course RT 2) Pre-op CRT 3) Pre-op long course RT plus postop CTX 4) Pre-op CRT plus postop CTX 5 year follow-up	DRE, CT +/- ERUS	17.1* (9.6 with postop CTX)	8.7* (7.6 with postop CTX)	65.8	64.8

Pre-op = preoperative; post-op = postoperative

CRT = chemoradiation therapy; CTX = chemotherapy; RT = radiation therapy

CT = computed tomography; DRE = digital rectal examination; ERUS = endorectal ultrasound;

MRI = magnetic resonance imaging

* denotes $p<0.05$

Table A.3 Summary of previous meta-analyses of diagnostic accuracy of MRI for staging rectal cancer

Meta-analysis	Databases searched	Studies	Modalities	Pre-operative RT or CRT included?	Statistical analysis	Staging elements	Results	Summary	Limitations
Kwok (2000)	Medline	83 studies (1981-1998)	MRI, ERUS, CT	Yes	Descriptive (Simple pooled)	T-category	MRI Sens 86%, Spec 77% ERUS Sens 93%,Spec 78% CT Sens 78%, Spec 63%	MRI with endorectal coil most accurate for both T-category and nodes	1) Older studies, included outdated technology 2) Analysis descriptive, did not account for heterogeneity 3) No formal study quality assessment 4) Did not examine CRM
						Nodes	MRI Sens 65%, Spec 80% ERUS Sens 71%, Spec 76% CT Sens 52%, Spec 78%		
Bipat (2004)	Medline, EMBASE, Cochrane, CANCERLIT	90 studies (1985-2002)	MRI, ERUS, CT	No	Bivariate random effects	T-category	MRI Sens 82% (74-87), Spec 76% (65-84) ERUS Sens 90% (88-92), Spec 75% (69-81) CT Sens 79% (74-84), Spec 78% (73-83)	MRI less accurate than ERUS for T-category, equivalent for nodes	1) Older studies, included outdated technology 2) Did not examine CRM 3) No formal study quality assessment

Table A.3 (continued)

Meta-analysis	Databases searched	Studies	Modalities	Pre-operative RT or CRT included?	Statistical analysis	Staging elements	Results	Summary	Limitations
						Nodes	<u>MRI</u> Sens 66% (54-76) Spec 76% (59-87) <u>ERUS</u> Sens 67% (60-73), Spec 78% (71-84) <u>CT</u> Sens 74% (67-80), Spec 55% (43-67)		
Lahaye (2005)	Medline	82 studies (1985-2005)	MRI, ERUS, CT	Yes	Moses-Littenberg fixed effects model	CRM	<u>MRI</u> Sens 80%, Spec 80%	MRI best modality for predicting CRM status; all modalities equivalent for nodes	1) Statistical analysis inadequate to account for heterogeneity 2) No separate analysis by coil type or radiotherapy status 3) No formal study quality assessment
						Nodes	<u>MRI</u> DOR 6.53 <u>ERUS</u> DOR 8.83 <u>CT</u> 5.86		

Table A.3 (continued)

Meta-analysis	Databases searched	Studies	Modalities	Pre-operative RT or CRT included?	Statistical analysis	Staging elements	Results	Summary	Limitations
Purkayastha (2007)	Medline, EMBASE, Cochrane, Ovid	9 studies (2001-2005)	MRI	Yes	Fixed effects, pooled	CRM	MRI Sens 94% (90-97), Spec 85% (81-89) DOR 57.21 (18.21-179.77)	MRI has good accuracy for CRM	1) Statistical analysis inadequate to account for heterogeneity 2) No separate analysis by radiotherapy status

CT = computed tomography; ERUS = endorectal ultrasound; MRI = magnetic resonance imaging
CI = confidence interval; DOR = diagnostic odds ratio; Sens = sensitivity; Spec = specificity
CRT = chemoradiotherapy; RT = radiotherapy

Table A.4 Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool

Item		Yes	No	Unclear
1.	Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	( )	( )
2.	Were selection criteria clearly described?	( )	( )	( )
3.	Is the reference standard likely to correctly classify the target condition?	( )	( )	( )
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	( )
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	( )	( )	( )
6.	Did patients receive the same reference standard regardless of the index test result?	( )	( )	( )
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	( )	( )	( )
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?	( )	( )	( )
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	( )	( )
10.	Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	( )
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	( )
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	( )	( )	( )
13.	Were uninterpretable/ intermediate test results reported?	( )	( )	( )
14.	Were withdrawals from the study explained?	( )	( )	( )

Table A.5 Modified QUADAS tool with quality assessment criteria

	Item	Criteria for fulfilling item
1	Was the spectrum of patients representative of the patients who will receive the test in practice?	Tumours from all levels of rectum (upper, mid, low) Each of T2 and T3 tumours must make up at least 25% of the sample[1]
2	Is the reference standard likely to classify the target condition correctly?	Reference standard is complete histopathology
3	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	Time period ≤8 weeks between MRI and surgery (≤2 weeks if patient received preoperative short-course radiotherapy after MRI acquisition)
4	Did the whole sample or a random selection of the sample, receive verification using the intended reference standard?	Complete histopathology must be available for all patients
5	Did patients receive the same reference standard irrespective of the index test result?	Complete histopathology must be available for all patients
6	Was the reference standard independent of the index test?	All studies fulfill this item (histopathology is independent of MRI)
7	Were the reference standard results interpreted without knowledge of the results of the index test?	Explicit statement that pathologist assessed specimens without knowledge of MRI findings
8	Were the index test results interpreted without knowledge of the results of the reference standard?	Radiologist(s) assessed images without knowledge of histopathology findings; this item was automatically fulfilled for all prospective studies (as the MRI must have been reviewed before histopathology results were available) but required an explicit statement to be considered fulfilled by retrospective studies
9	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	Radiologist(s) was/were aware of the purpose of the MRI (rectal cancer staging) but was/were otherwise blinded to other clinical data
10	Were uninterpretable/intermediate test results reported?	All patients accounted for
11	Were withdrawals from the study explained?	Withdrawals from the study were explained
12	Did the execution of the index test meet currently accepted standards?	T2 weighted, high resolution axial images performed perpendicular to long axis of tumour-bearing rectum
13	Did the execution of the reference standard meet currently accepted standards?	Multi-slice "Quirke" method used to assess histopathology specimen

[1]Rationale: sufficient proportions of T2 and T3 tumours must be present to represent their true prevalence and adequately test observors' ability to distinguish between them

Figure A.1 Medline Search Strategy

	MRI & Related Terms Component
1	exp magnetic resonance imaging/
2	nmr imaging.mp.
3	zeugmatograph*.mp.
4	mr tomograph*.mp.
5	nmr tomograph*.mp.
6	proton spin* tomograph*.mp.
7	magneti#ation transfer contrast imag*.mp.
8	(mri adj2 scan*).mp.
9	chemical shift* imag*.mp.
10	(magnetic resonance adj2 imag*).mp.
11	(MR adj2 imag*).mp.
12	(NMR adj2 imag*).mp.
13	(diffusion weighted adj2 imag*).mp.
14	(T1-weighted adj2 imag*).mp.
15	(T2-weighted adj2 imag*).mp.
16	mri.mp.
17	dwi.mp.
18	dwi.tw.
19	magnetic resonance spectroscop*.mp.
20	MRS.tw.
21	(dynamic contrast-enhanc* adj2 (imag* or MR or MRI)).mp.
22	"3.0 tesla".mp.
23	rectal coil*.mp.
24	(endorectal adj2 coil*).mp.
25	(endo-rectal adj2 coil*).mp.
26	gadolidium.mp.
27	gadolinium.mp.
28	or/1-27
	Rectal Cancer & Related Terms Component
29	exp rectal neoplasms/
30	(Adenocarcinom: adj3 (rect: or mesorectal* or endorectal* or extramesorectal* or rectosigmoid*)).mp.
31	(Cancer: adj3 (rect: or mesorectal* or endorectal* or extramesorectal* or rectosigmoid*)).mp.

32	(Carcin: adj3 (rect: or mesorectal* or endorectal* or extramesorectal* or rectosigmoid*)).mp.
33	(Neoplas: adj3 (rect: or mesorectal* or endorectal* or extramesorectal* or rectosigmoid*)).mp.
34	(Tumor: adj3 (rect: or mesorectal* or endorectal* or extramesorectal* or rectosigmoid*)).mp.
35	(Tumour: adj3 (rect: or mesorectal* or endorectal* or extramesorectal* or rectosigmoid*)).mp.
36	(Adenom: adj3 (rect: or mesorectal* or endorectal* or extramesorectal* or rectosigmoid*)).mp.
37	(Malignan: adj3 (rect: or mesorectal* or endorectal* or extramesorectal* or rectosigmoid*)).mp.
38	or/29-37
	Cancer Staging & Related Terms Component
39	Neoplasm Staging/
40	(stage or stages or staged or staging).mp.
41	(restage or re-stage or restages or re-stages or restaged or re-staged or restaging or re-staging).mp.
42	(duke or dukes).mp.
43	ajcc.mp.
44	tumo?r-node?-metastasis.mp.
45	tnm.mp.
46	circumferential resection margin?.mp.
47	mesorectal fascia.mp.
48	meso-rectal fascia.mp.
49	radial resection margin?.mp.
50	resection margin?.mp.
51	or/39-50
52	**28 and 38 and 51 → MRI + Rectal Cancer + Cancer Staging →**
	HIRU (McMaster) "Diagnosis" search filter
53	evaluation studies.pt.
54	(evaluation adj1 (study or studies)).mp.
55	validation studies.pt.
56	(validation adj1 (study or studies)).mp.
57	exp "sensitivity and specificity"/
58	exp diagnostic errors/
59	likelihood functions/
60	(likelihood: or likelihood ratio:).mp.
61	or/53-60
	Radiology "Diagnosis" search filter
62	exp "sensitivity and specificity"/
63	False positive reactions/
64	False negative reactions/

65	di.fs.
66	sensitivity.tw.
67	(predictive adj4 value*).tw.
68	distinguish*.tw.
69	differentiat*.tw.
70	enhancement.tw.
71	identif*.tw.
72	detect*.tw.
73	diagnos*.tw.
74	accura*.tw.
75	comparison.tw.
76	or/62-75
77	76 or 61 → HIRU or Radiology search filters
78	52 and 77 → (MRI + Rectal CA + Cancer Staging) + (HIRU or Radiology search filters) →
	Limits applied:
79	limit 78 to yr="2000 -Current"
80	limit 79 to english language
81	**remove duplicates from 80 → final result →**