

Clinical Radiology

3-Tesla Magnetic Resonance Imaging and Computed Tomography accurately identify T3c disease in colon cancer --Manuscript Draft--

Manuscript Number:	CRAD-D-16-00591R1
Full Title:	3-Tesla Magnetic Resonance Imaging and Computed Tomography accurately identify T3c disease in colon cancer
Article Type:	Original Paper
Corresponding Author:	Chris Hunter, MBChB Royal Marsden NHS Foundation Trust Sutton, - - UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Royal Marsden NHS Foundation Trust
Corresponding Author's Secondary Institution:	
First Author:	Chris Hunter, MBChB
First Author Secondary Information:	
Order of Authors:	Chris Hunter, MBChB
	Muhammed Siddiqui, MBChB
	Theo Georgiou Delisle, MBBS
	Helena Blake, MS
	Nelesh Jeyadevan, MS
	Muti Abulafi, MS
	Ian Swift, MB BS
	Paul Toomey, MD
	Gina Brown, MD
Order of Authors Secondary Information:	
Abstract:	Aim: To compare the preoperative staging accuracy of CT and 3T MRI in colon cancer, and to investigate the prognostic significance of identified risk factors. Materials and methods: 58 patients undergoing primary resection of their colon cancer were prospectively recruited, with 53 patients included for final analysis. Accuracy of CT and MRI were compared for two readers, using postoperative histology as the gold standard. Patients were followed-up for a median of 39 months. Risk factors were compared by modality and reader in terms of metachronous metastases and DFS, stratified for adjuvant chemotherapy. Results: Accuracy for the identification of T3c+ disease was non-significantly greater on MRI (75% and 79%) than CT (70% and 77%). Differences in the accuracy of MRI and CT for identification of T3+ disease (MRI 75% and 57%, CT 72% and 66%) and N+ disease (MRI 62% and 63%, CT 62% and 56%) were also non-significant. Identification of EMVI+ disease was significantly greater on MRI (75% and 75%) than CT (79% and 54%) for one reader (p=0.029). T3c+ disease on histopathology was the only risk factor which demonstrated a significant difference in rate of metachronous metastases (OR 8.6, p=0.0044) and DFS stratified for adjuvant therapy (OR 4, p=0.048). Conclusion: T3c or greater disease is the strongest risk factor for predicting DFS in colon cancer,

	and is accurately identified on imaging. T3c+ disease may therefore be the best imaging entry criteria for trials of neoadjuvant treatment.
--	---

Title

3-Tesla Magnetic Resonance Imaging and Computed Tomography accurately identify T3c disease in colon cancer.

Authors

Chris Hunter

Muhammed Siddiqui

Theo Georgiou Delisle

Helena Blake

Nelesh Jeyadevan

Muti Abulafi

Ian Swift

Paul Toomey

Gina Brown

Institution

The Royal Marsden NHS Foundation Trust,

Downs Road,

Sutton,

SM2 5PT

United Kingdom.

Corresponding Author

Chris Hunter, Tel 0208 915 6067, Fax 0208 915 6721, Email chris.hunter2@nhs.net

Grants

This research project was supported by the Bowel Diseases Research Foundation grant; Croydon University Hospital; the NIHR Biomedical Research Centre at the Royal Marsden NHS Foundation Trust; the Wellcome Trust. The funding sources did not have any input into study design or reporting.

Acknowledgements

The authors would like to the MR radiographers at The Royal Marsden Hospital and Croydon University Hospital for their help in running this research project.

Guarantor of the integrity of the entire study

Gina Brown

Study concepts and design

Gina Brown, Chris Hunter, Muti Abulafi, Ian Swift, Paul Toomey, Helena Blake, Nelesh Jeyadevan

Literature research

Chris Hunter

Clinical studies

Chris Hunter, Muhammed Siddiqui, Theo Georgiou Delisle

Experimental studies/data analysis

Gina Brown, Helena Blake, Nelesh Jeyadevan

Statistical analysis

Gina Brown, Chris Hunter

Manuscript preparation

Gina Brown, Chris Hunter

Manuscript editing

Chris Hunter, Muhammed Siddiqui, Theo Georgiou Delisle, Helena Blake, Nelesh Jeyadevan, Muti Abulafi, Ian Swift, Paul Toomey, Gina Brown

Editor comments:

- 1) Title: Amended to remove “ which strongly predicts disease free survival”.
- 2) Clarification of the original reason for this research: Introduction (lines 28-32) amended to clarify the primary and secondary aims of this study and the underlying hypothesis.
- 3) This suggests it was expected that MRI would give far better accuracy for staging than CT but this was disproved in the analysis. This itself is quite a useful finding - what this study actually seems to show is that in fact CT is adequate for staging and there is no need to perform an additional MRI in these patients with the attendant extra costs/time/patient discomfort that this entails: Sentence added to conclusion (line 285-287), highlighting that CT provides adequate local staging.
- 4) Was 3T MRI thought likely to be better than 1.5T MRI in these patients: Sentence added to conclusion stating that accuracy with 1.5T MRI and 3T MRI was similar (313-314).
- 5) Why were the MR sequences chosen: Sentence added to conclusion (311-315) explaining that the choice of sequences was based on a small pilot study, but that we feel inclusion of T1 weighted images and DWI may have been detrimental.
- 6) The selection and labelling of the images could be improved: I have updated the images and legends.

Reviewer #1:

- 1) how sensitivity can be applied to T staging: (clarified that this refers to T1-2 vs T3-4) (lines 22-24 and 26)
- 2) Reference 6 did include 21 studies, but only 19 were included for the assessment of T-staging accuracy.
- 3) Line 37 - ethical approval - was this granted prior to the ethical outcome becoming favourable opinion, rather than outcome? – I’m sorry, I don’t understand this query.
- 4) Would like to know more about the follow up and breakdown of this if possible - it is stated that median is 39/12, with a range of 1 - 56/12 - one month is clearly very short follow up - should this not be excluded, and have a minimum length of f/up? Given that recruitment finished in 11/2012, is there an option to extend follow up to now, so that all have a minimum length of > 3 - 4 years?: All patients were followed up from recruitment until May 2016. However, some patients were lost to follow up and in this case they were censored at the date of their last follow up. I have added a sentence to the methods to clarify this (line 123)
- 5) few American spellings: These have been changed to English versions
- 6) lines 77/114 - should be years rather than years’: This has been changed
- 7) line 115 - RC Pathologists: This has been changed
- 8) Should Bracco etc be in parentheses at the start?: This has been changed
- 9) 307 - should be Radiologists: This has been changed

Reviewer #2:

- 1) A reference for the meta-analysis referred to in line 22-24 is required: I have included this reference
- 2) Line 102 - I believe the 10mm cut off for size assessment of lymph nodes is a short axis measurement, but this should be made clear in the text: I have amended this (line 104)
- 3) T1 – unfortunately it has not been possible to include an example of a T1 tumour. Only 2 T1 tumours were included in the study and neither demonstrates well the difference between a T1 and T2 tumour.
- 4) Figure 1 – This has been altered to include sagittal images to illustrate the tumour more clearly.
- 5) Figure 4 – This has been amended to include coronal and sagittal MRI images of a lymph node deposit, which more clearly demonstrates irregular borders and signal inhomogeneity.
- 6) Grammatical and spelling errors – These have been corrected.

Aim:

To compare the preoperative staging accuracy of CT and 3T MRI in colon cancer, and to investigate the prognostic significance of identified risk factors.

Materials and methods:

58 patients undergoing primary resection of their colon cancer were prospectively recruited, with 53 patients included for final analysis. Accuracy of CT and MRI were compared for two readers, using postoperative histology as the gold standard. Patients were followed-up for a median of 39 months. Risk factors were compared by modality and reader in terms of metachronous metastases and DFS, stratified for adjuvant chemotherapy.

Results:

Accuracy for the identification of T3c+ disease was non-significantly greater on MRI (75% and 79%) than CT (70% and 77%). Differences in the accuracy of MRI and CT for identification of T3+ disease (MRI 75% and 57%, CT 72% and 66%) and N+ disease (MRI 62% and 63%, CT 62% and 56%) were also non-significant. Identification of EMVI+ disease was significantly greater on MRI (75% and 75%) than CT (79% and 54%) for one reader ($p=0.029$). T3c+ disease on histopathology was the only risk factor which demonstrated a significant difference in rate of metachronous metastases (OR 8.6, $p=0.0044$) and DFS stratified for adjuvant therapy (OR 4, $p=0.048$).

Conclusion:

T3c or greater disease is the strongest risk factor for predicting DFS in colon cancer, and is accurately identified on imaging. T3c+ disease may therefore be the best imaging entry criteria for trials of neoadjuvant treatment.

1 Title:

2 3-Tesla Magnetic Resonance Imaging and Computed Tomography accurately identify T3c disease in

3 colon cancer

4 Running head:

5 3T MRI and CT in the preoperative staging of colon cancer

6

7 Introduction

Accurate preoperative staging in colon cancer is increasingly important. The FOXTROT trial is currently investigating pre-operative chemotherapy in locally advanced colon cancer, following encouraging pathological responses seen in the pilot phase of the study (1). Selection of patients who may benefit from this approach requires accurate preoperative local staging.

The proportion of patients undergoing laparoscopic resection of colon cancers is increasing steadily, with 40% of colorectal procedures performed laparoscopically in England in 2012 compared to less than 5% in 2005/6 (2). Patients with large or locally advanced colon cancers may not be appropriate for laparoscopic surgery; the COLOR trial report recommends pre-operative assessment to identify patients with locally advanced disease who are not suitable for laparoscopic surgery (3).

Preoperative imaging must, therefore, identify not only patients with metastatic disease who may not benefit from operative treatment, but also those with early stage disease who will not benefit from neoadjuvant therapy, and those with adjacent organ involvement who need an open surgical approach or more radical surgery.

Contrast-enhanced CT of the chest, abdomen and pelvis is currently recommended for the local staging of colon cancer and identification of distant metastases (4). A meta-analysis of eleven studies reported that the sensitivity of CT for T-staging (T1-2/T3-2) in colon cancer was relatively high at 77%, but that specificity was surprisingly low at 3% due mainly to underestimating T-stage (5).

MRI is the recommended modality for the local staging of rectal cancer (4) and achieves accurate local staging of rectal cancers. On meta-analysis of 19 studies for T-staging assessment (T1-2/T3-4), sensitivity and specificity were 87% and 75% respectively (6).

The primary aim of this study was to compare the accuracy of local staging for colon cancers using 3T MRI and CT scans using histology as the gold standard. We hypothesized that MRI would provide more accurate local staging than CT. A secondary aim of this study was to investigate directly the

prognostic value of adverse risk features for predicting disease recurrence and disease-free survival (DFS).

Methods

Patients

This was a prospective study of diagnostic accuracy, comparing 3T MRI and CT using histological assessment as the gold standard. This study also assessed the accuracy of 1.5T MRI, and these results have been reported separately (7). Data was collected prospectively. The study was designed according to the STARD guidelines (8).

Ethical approval was obtained from the National Research Ethics Service, before commencing recruitment (REC reference number 10/H0801/21). The study was conducted by the conditions of ethical approval. Informed consent was obtained from all participants before enrolment. The trial was registered with ClinicalTrials.gov (registration number NCT01187641).

Consecutive patients over the age of 18 presenting with adenocarcinoma of the colon to three institutions, who were undergoing primary surgery to excise their colon cancer, were included in the study. Patients were excluded if they had a contraindication to MRI scanning or the administration of gadolinium.

Patients were withdrawn from the study if metastatic disease was identified on their preoperative MRI scan which required neoadjuvant treatment, or if they were unable to tolerate MRI scanning.

Procedures

Patients enrolled in the study underwent a preoperative MRI scan of their abdomen and pelvis, in addition to their standard preoperative CT scan, which was performed on a 3.0T MRI scanner (Achieva 3.0T, Philips Healthcare, Best, Netherlands). Patients fasted for 4 hours before MRI scanning. Bowel cleansing preparation was not performed. Twenty milligrammes of intramuscular

hyoscine butyl bromide (Buscopan, Boehringer Ingelheim Limited, Bracknell, United Kingdom) was administered. One litre of orally administered contrast was used to fill the small bowel - 225 ml 4.9% weight/volume barium sulphate diluted in water to 1L, (E-Z-CAT, Bracco UK Ltd, High Wycombe, United Kingdom). All images were acquired using a SENSE XL Torso coil.

A breath-hold Fully Rewound Gradient Echo sequence (FRGE) coronal scout scan was taken to identify the tumour and select the target volume. A respiratory triggered high-resolution Turbo Spin Echo (TSE) sequence was acquired perpendicular to the bowel wall through the tumour (835-1454/80 ms TR/TE). The voxel size in the high-resolution images was 0.73-0.99 x 0.73-0.99 x 2.0 mm with a field of view of 350-400 mm. Slice thickness was 3mm. Where possible within the time constraints of the scan, additional T2 weighted images were taken in complimentary planes.

Diffusion-weighted imaging (DWI) was then acquired through the tumour and mesentery, using a spin-echo sequence with echoplanar imaging and spectral attenuated inversion recovery fat suppression. Repetition time was 4800ms and echo time was 68ms. Four b-values were used: 0, 100, 500 and 750 s/mm².

Following the administration of IV gadolinium, dynamically enhanced fat suppressed T1 weighted spoiled gradient echo sequences were acquired. Total imaging time was 30 to 45 minutes.

The preoperative CT scans used in the trial were the standard staging CT scans obtained for clinical assessment for each patient at each of the participating institutions. The CT scans were performed on one of two scanners: Siemens SOMATON Sensation 16 (Siemens Healthcare, Erlangen, Germany) or GE LightSpeed VCT (GE Healthcare, Buckinghamshire, UK). Imaging was performed with intravenous contrast in the portal venous phase. Images were reconstructed with a slice thickness of 2mm or 5mm and an in-plane voxel size of 0.61x0.61mm to 0.85x0.85mm. Multiplanar reformatting was available.

Evaluation

Each CT scan and 3.0T MRI scan was reported independently by each of two consultant gastrointestinal radiologists, with 10 - 15 years experience of reporting GI cancers. Each observer reported each imaging study in a randomized order on a separate occasion more than four weeks apart.

T1 tumors were defined as those limited to the submucosa, identified on MRI by abnormal intermediate signal which did not extend into the circular muscle layer and on CT by the projection of a colonic lesion without any visible distortion of the bowel wall layers. T2 tumours were limited to the muscularis propria, identified on MRI by abnormal intermediate signal intensity extending into the muscularis propria and on CT by asymmetrical thickening of the colon wall projecting intraluminally with preservation of smooth muscle coat and clear adjacent pericolic fat. An example is shown in Figure 1.

T3 tumours were limited to the subserosa or peri-colic fat, identified on MRI by a broad based bulge or nodular projection of intermediate tumour intensity beyond the outer muscle coat and on CT by smooth or nodular extension of a discrete mass of tumour tissue beyond the contour of the bowel wall. An example is shown in Figure 2.

T4 tumours were defined as those that invade the serosa or adjacent organs, identified on MRI by extension of the intermediate tumour signal into adjacent organs or through the peritoneal surface, and on CT by the presence of nodular penetration of the tumour through the peritonealised areas of the muscle coat or an advancing edge of the tumour penetrating adjacent organs. An example is shown in Figure 3. The MRI criteria were adapted from those used in the evaluation of rectal cancer (11), and the CT criteria were adopted from established criteria for the interpretation of CT images of colon cancer (12).

Lymph nodes were assessed on MRI based on their border and homogeneity of signal intensity, which has been shown to have greater specificity for the identification of metastatic lymph nodes in rectal cancer than the use of size criteria alone (13, 14). Lymph nodes were considered to be involved if they had an irregular border or inhomogeneous signal intensity, assessed on the T2 weighted TSE imaging. An example is shown in Figure 4. On CT, any nodal mass larger than 1 cm in the short axis diameter or a group of three or more nodes were considered metastatic, as these are the established criteria for assessing lymph node involvement on CT (12).

Extramural vascular invasion (EMVI) was identified on MRI by the intermediate signal intensity growing into or along a vessel. Blood vessels were identified on T2-weighted MRI imaging by a linear signal void in continuity on adjacent slices (15). EMVI was identified on CT by the irregular expansion of peritumoural veins (16). An example is shown in Figure 5. EMVI was reported as small, medium, or large vessel EMVI on MRI and CT. Any size of EMVI identified on imaging was considered positive for comparison with histopathology, as EMVI is reported as positive or negative on histopathology.

Reference standard

Histopathological assessment of the excised colon cancer specimen was used as the gold standard against which imaging was compared. Histopathological assessment was carried out by one of two consultant gastrointestinal pathologists with more than ten years experience in reporting colon cancer. Histopathology was prospectively reported according to the Royal College of Pathologists minimum dataset (17).

In addition to assessing the accuracy of CT and MRI against histopathology, the prognostic significance of each end point was assessed in predicting metachronous metastases and 3-year disease free survival (DFS). Disease status was evaluated on clinical follow up with 6-monthly serum CEA levels, annual CT thorax, abdomen and pelvis and colonoscopy at one and five years. Where patients were lost to follow up, they were censored at the date of their last follow up.

Endpoints

The primary endpoint was the assessment of advanced T-stage (T3+) tumours, which is the current radiological entry criteria for the FOXTROT trial for patients who are of younger age and in general good health (18). Stage-for-stage T-staging accuracy is also reported.

T-stage and depth of extramural invasion were also combined to identify poor prognosis tumours (T3 tumours with more than 5mm of extramural invasion and T4 tumours, T3c+). This distinction has been demonstrated to have greater prognostic significance than the distinction between T2 and T3 tumours (19), and it is the current entry criteria for the FOXTROT trial for patients who are not of younger age or in general good health (18).

Lymph node metastases are reported as lymph node positive (N1 and N2 tumours, N+) and lymph node negative (N0 tumours). Nodal positivity is the most clinically significant endpoint identified on histopathology, with lymph node positive (Stage III) deriving the greatest benefit from adjuvant chemotherapy in clinical trials (20).

EMVI is assessed as EMVI positive (EMVI+). MRI detected EMVI in rectal cancer has been demonstrated to be associated with significantly worse 3-year recurrence free survival, with a prognosis similar to histopathologically detected EMVI (15).

Sample size calculation

The sample size was based on the primary endpoint, which was a comparison of the accuracy of MRI and CT in identifying T3+ disease using histology as the gold standard. We anticipated an accuracy of 0.70 and 0.90 for CT and MRI, respectively. Since MRI and CT were performed for each patient, a paired design was used for determining sample size, with a two-tailed alternative hypothesis. Given a significance level, α , of 5% and a power of 80%, the minimum number of patients required was 62 patients.

Statistical analysis

Patient characteristics are presented as means with standard deviation (SD) for normally distributed quantitative data, medians with interquartile ranges (IQR) for non-normally distributed quantitative data, and frequencies with percentages for qualitative data.

For ordinal outcomes the accuracy of MRI and CT in assessing T-stage is reported for each reader.

For categorical outcomes the accuracy, sensitivity and specificity are reported with 95% CIs. The significance of the difference in accuracy between CT and MRI is assessed using McNemar's test.

Interobserver agreement for the two radiologists reporting CT and MRI is determined using Cohen's kappa (23). The magnitude of the agreement is quantified as follows: a kappa of <0.2 denotes a slight agreement; 0.2 to 0.4, a fair agreement; 0.4 to 0.6, a moderate agreement; 0.6 to 0.8, a substantial agreement, and >0.8 , an almost perfect agreement.

The incidence of metastatic disease during follow-up according to risk factors was compared using Fisher's exact test. DFS curves were estimated using the Kaplan-Meier method. Multivariable analysis was performed using Cox proportional hazards model, stratified according to whether patients had received adjuvant chemotherapy.

Analyses were performed using R-3.2.1 (R Core Team, Vienna, Austria) (24), including psych (25) and survival (26) packages. All evaluable subjects are included in the analyses.

Results

Recruitment

Recruitment commenced November 2010 and finished in November 2012. A total of 58 patients were recruited, from a total of 158 eligible patients (37%). Patients were not recruited because they did not consent to the trial (41 patients), because they could not be contacted with sufficient time to

participate (35 patients), because it was not possible to arrange the MRI scan (16 patients), and for other reasons (6 patients).

Of the 58 patients recruited into the study, five patients (9%) were withdrawn. One patient failed to attend his MRI appointment, one patient had no tumour identified on histopathological assessment, and three patients did not undergo surgery.

Fifty-three patients are included in the final analysis. One patient (1.9%) had an unresectable tumour due to duodenal invasion identified at surgery (T4) and is included for the assessment of T-stage, but not for other endpoints.

Patient and tumor characteristics

The mean age of patients recruited into the study was 69.3 (SD $\pm$ 13.6) years. Thirty-four patients (64%) were male. The median delay from MRI scan to surgery was nine days (IQR 6-14 days), compared to 40 days (IQR 29-50 days) from CT to surgery. This difference is significant ($p < 0.001$, Wilcoxon signed rank test). The reason for this difference is that all CT scans were necessarily performed before recruitment into the study, as patients had to be discussed in the MDT and considered suitable for surgery to be eligible for recruitment. However, there was no evidence of tumour progression when we compared the overall tumour staging on CT and MRI (median T-stage T3 on CT and MRI, median N-stage N1 on CT compared to N0 on MRI, proportion EMVI positive on CT 38.7% compared to 30.2% on MRI).

Eight tumors were located in the caecum (15.1%), thirteen in the ascending colon (24.5%), five in the hepatic flexure (9.4%), six in the transverse colon (11.3%), four in the descending colon (7.5%), and seventeen in the sigmoid colon (32.1%).

The pathological stage of tumours was T1 in two patients (3.8%), T2 in nine patients (17.0%), T3 in twenty-eight patients (52.8%), and T4 in fourteen patients (26.4%). The majority of patients

recruited to the trial (30 patients, 56.6%) had N0 disease on pathological staging, followed by N1 disease (18 patients, 34.0%) and N2 disease (4 patients, 7.5%).

The fifty-two patients who underwent potentially curative resection for their colon cancer were included in the survival analysis. Median follow-up was 39 months (range 1 to 56 months). Twenty-one patients (40%) received adjuvant chemotherapy. Twelve patients (23%) developed metastatic disease during the follow-up period, 8/21 (38%) in the chemotherapy group and 4/31 (13%) in the group who did not receive adjuvant chemotherapy. Overall 3-year DFS for the whole group was 68.9% (95% CI 56.5%-84.0%).

MRI accuracy

The overall accuracy using MRI for the identification of T3+ disease was 75% (95% CI 62-85%) for Reader 1 (R1) and 57% (95% CI 43-69%) for Reader 2 (R2). Accuracy, sensitivity and specificity for identifying T3 and greater disease is shown in Table 1. Stage-for-stage T-stage accuracy was 55% (95% CI 41-67%) for R1 and 40% (95% CI 28-53%) for R2. Cohen's weighted kappa for the inter-observer agreement between R1 and R2 was 0.44 (0.27-0.61) indicating moderate agreement.

Accuracy for the identification of T3c+ disease was 75% (95% CI 62-85%) for R1 and 79% (95% CI 67-88%) for R2. Accuracy, sensitivity and specificity are shown in Table 2. Cohen's kappa for interobserver reliability between R1 and R2 was 0.51 (95% CI 0.29 to 0.73), indicating moderate agreement.

Accuracy for the identification of N+ disease was 62% (95% CI 48-74%) for R1 and 63% (95% CI 50-75%) for R2. Accuracy, sensitivity and specificity are shown in Table 3. Cohen's kappa for interobserver reliability between R1 and R2 was 0.28 (95% CI 0.04 to 0.55) indicating fair agreement.

Accuracy for the identification of EMVI+ disease was 75% (95% CI 62-85%) for both R1 and R2. Accuracy, sensitivity and specificity are shown in Table 4. Cohen's kappa for interobserver reliability between R1 and R2 was 0.34 (95% CI 0.12 to 0.56) indicating fair agreement.

CT accuracy

The overall accuracy using CT for the identification of T3+ disease was 72% (95% CI 58-82%) for R1 and 66% (95% CI 53-77%) for R2. Accuracy, sensitivity and specificity are shown in Table 1. The CT staging accuracy for stage-for-stage T-staging was 45% (33-59%) for R1 and 47% (34-60%) for R2. Cohen's weighted kappa was 0.59 (0.39-0.79) for the agreement between R1 and R2, indicating moderate agreement.

Accuracy for the identification of T3c+ disease was 70% (95% CI 56-80%) for R1 and 77% (95% CI 64-87%) for R2. Accuracy, sensitivity and specificity are shown in Table 2. Cohen's kappa for interobserver reliability between R1 and R2 was 0.45 (95% CI 0.22-0.68) indicating moderate agreement.

Accuracy for the identification of N+ disease was 62% (95% CI 48-74%) for R1 and 56% (95% CI 42-68%) for R2. Accuracy, sensitivity and specificity are shown in Table 3. Cohen's kappa for interobserver reliability between R1 and R2 was 0.43 (95% CI 0.2 to 0.67) indicating moderate agreement.

Accuracy for the identification of EMVI+ disease was 79% (95% CI 66-88%) for R1 and 54% (95% CI 41-67%) for R2. Accuracy, sensitivity and specificity are shown in Table 4. Cohen's kappa for interobserver reliability between R1 and R2 was 0.18 (95% CI -0.08 to 0.44) indicating slight agreement.

Difference in accuracy between MRI and CT

The primary end-point for this study was the difference in accuracy of CT and MRI in the identification of T3+ disease. MRI was slightly more accurate than CT for R1 (75% vs. 72%), and slightly less accurate for R2 (57% vs. 66%). These differences were not significant ($p=0.75$ and $p=0.33$ respectively, McNemar's test).

The difference in accuracy between CT and MRI in assessing T3c+ disease and N1+ disease was also non-significant for both readers ($p=0.51$ to 1.0). The difference in accuracy between CT and MRI for assessing EMVI+ disease was also non-significant for R1 ($p=0.79$). R2 was significantly more accurate assessing EMVI using MRI (75%) than CT (54%), ($p=0.029$).

Predicting metastases and disease-free survival

The odds ratios for developing metastatic disease during follow-up for each risk factor, assessed on CT, MRI and histopathology, are shown in Table 5. The strongest predictor of subsequent disease recurrence was T3c+ disease identified on histopathology, with an odds ratio of 8.6 for developing metastatic disease during follow-up. This was highly significant ($p=0.0044$, Fisher's exact test). T3c+ disease identified on CT by R2 also approached statistical significance ($p=0.081$, OR 3.4). Other risk factors on imaging and histopathology did not demonstrate a significant difference in rates of metastatic disease during follow-up in this small group.

Adjuvant chemotherapy was also associated with a significantly increased risk of metastatic disease (OR 4, $p=0.048$). This is because adjuvant chemotherapy was selectively administered to patients with adverse risk features on histopathological assessment, principally lymph node positivity.

The three most influential adverse risk features (T3c+, N+ and EMVI+) were therefore assessed in multivariate analysis according to DFS using Cox proportional hazards model for each imaging modality/reader, with patients stratified according to whether or not they had received chemotherapy (Table 6). T3c+ disease on histopathology remained a statistically significant risk factor for poorer DFS (hazard ratio 8.6 ($p=0.0044$)). T3c+ disease identified on MRI by R2 also approached statistical significance (hazard ratio 3.6, $p=0.056$). Other risk factors were not significant in this group.

Discussion

The only significant predictor of DFS in this study was T3c or greater disease, assessed on histopathology. In multivariate analysis, stratified according to patients who had received adjuvant chemotherapy, patients with T3c+ disease had a hazard ratio of 8.6 for developing metastatic disease. This is in keeping with previous pathological studies in rectal cancer which have demonstrated significantly higher recurrence rates and poorer DFS for T3 tumours with 5mm or more extramural spread, and similar survival for T2 tumours and T3 tumours with less than 5mm extramural spread (19). This study confirms that this risk factor is also significant in colon cancer.

The FOXTROT study used this inclusion criterion (T3 tumours with 5mm or more extramural spread on CT) in the pilot phase of the study (1). However, in the current trial protocol, the criteria have been extended to include all T3 tumours for young fit patients (18). Our results raise the possibility that the original criteria might be more likely to show a benefit for neoadjuvant chemotherapy.

The identification of T3c+ disease was also the most consistently accurate risk factor identified on imaging, with accuracy varying from 70% to 77% for CT and 75% to 79% for MRI. This also supports the recommendation that identification of T3c+ disease on imaging might be the best means of stratifying colon cancer patients into high and low-risk groups on pre-operative imaging.

Although the quality of CT scans achieved in this study was very consistent, there was more variation in the quality of the MRI scans. This was primarily due to difficulty controlling motion artefact, particularly when respiratory triggering was poor. When only the good quality MRI scans are considered (31/52, 60%), the accuracy of MRI in the identification of T3c+ disease was higher (81% for R1 and 77% for R2).

The primary aim of this project was to determine, on a per patient basis, whether the identification of T3 or greater disease is more accurate using MRI than using CT. There was no significant difference between the accuracy using CT and MRI for either reader. Accuracy was moderate on

both imaging modalities for both readers ranging from 57% for R2 on MRI to 75% for R1 on MRI, with accuracy on CT lying between these values (72% and 66% for R1 and two respectively). These results suggest that CT provides adequate local staging of colon cancer and that the additional cost of MRI for local staging is not currently justified.

The accuracy using CT for R1 (72%) and R2 (66%) are within the range of recent published results for the accuracy of CT in identifying T3 or greater disease; 76.2% to 83.3% (27), 70% to 82% (28), 62% to 80% (29), 88% (30).

Accuracy in the identification of EMVI was similar between modalities for R1 (79% on CT and 75% on MRI), although there was a slightly greater difference in accuracy between modalities for R2 (54% on CT and 75% on MRI). This was the only difference which achieved statistical significance ($p=0.029$ McNemar test) for either reader.

There is only one recent published study which investigates the accuracy of MRI in the staging of colon cancer (31). Rollven investigated the accuracy of 1.5T MRI and CT for the identification of locally advanced tumours (T3c+). Rollven reports accuracy using CT of 79.3% and 75.9% for two observers, which are comparable to the results of this study for accuracy using CT (70% for R1 and 77% for R2). Rollven reports accuracy using 1.5T MRI of 89.7% and 93.1%. These results represent a higher accuracy than we achieved in our study; 75% and 79% for two readers, although the 95% confidence intervals overlap.

There are several reasons why this study may have failed to show the level of accuracy for T-stage illustrated in the Rollven study. Both are small studies with wide confidence intervals for the reported accuracies. The current study is presented on an intention to treat basis for all recruited patients with evaluable results, with over 90% of recruited patients included for final evaluation. In the Rollven study, just fewer than 60% of recruited patients were included for final analysis. It is, therefore, possible that the Rollven study represents more optimal results in patients who can

tolerate a relatively lengthy MRI scan. Differences in the study protocols may also have influenced staging accuracy; the Rollven study only acquired respiratory triggered T2 weighted TSE sequences in multiple planes, where the acquisition of T1 and diffusion weighted imaging in our study restricted the acquisition of T2 weighted imaging in multiple planes. The choice of MRI sequences used in this study was based on a small pilot study in four healthy volunteers. However, the radiologists in our study found the T2 weighted TSE sequences were the most useful for staging, and acquisition of T1 weighted images and diffusion weighted images may have detracted from this study.

We had anticipated that 3T MRI might provide more accurate local staging than 1.5T MRI (results reported separately (7)). However, the accuracies achieved using 1.5T MRI and 3T MRI were similar.

There are several limitations to the current study. The study was underpowered due to slower than anticipated recruitment, there was a greater delay between CT and surgery than between MRI and surgery and the administration of adjuvant chemotherapy was dependent on histopathological staging.

In summary, this study demonstrates comparable accuracy in the local staging of colon cancer using CT and 3T MRI but fails to show the very high local staging accuracy using MRI reported in another recent study. This study supports the use of T3c+ disease as an imaging prognostic indicator when stratifying patients for neoadjuvant therapy.

References

1. Foxtrot Collaborative G. Feasibility of preoperative chemotherapy for locally advanced, operable colon cancer: the pilot phase of a randomised controlled trial. Lancet Oncol. 2012 Nov;13(11):1152-60.
2. LAPCO. National Training Programme for Laparoscopic Colorectal Surgery. 2012 [cited 2015 13/11/2015]; Available from: <http://lapco.nhs.uk/>.

- 332 3. Buunen M, Veldkamp R, Hop WC, Kuhry E, Jeekel J, Haglind E, et al. Survival after
333 laparoscopic surgery versus open surgery for colon cancer: long-term outcome of a randomised
334 clinical trial. *Lancet Oncol.* 2009 Jan;10(1):44-52.
- 335 4. Scottish Intercollegiate Guidelines Network (SIGN). Diagnosis and management of colorectal
336 cancer. Edinburgh: SIGN; 2011.
- 337 5. Leufkens AM, van den Bosch MA, van Leeuwen MS, Siersema PD. Diagnostic accuracy of
338 computed tomography for colon cancer staging: a systematic review. *Scand J Gastroenterol.* 2011
339 Jul;46(7-8):887-94.
- 340 6. Al-Sukhni E, Milot L, Fruitman M, Beyene J, Victor JC, Schmocker S, et al. Diagnostic accuracy
341 of MRI for assessment of T category, lymph node metastases, and circumferential resection margin
342 involvement in patients with rectal cancer: a systematic review and meta-analysis. *Ann Surg Oncol.*
343 2012 Jul;19(7):2212-23.
- 344 7. Hunter C, Blake H, Jeyadevan N, Abulafi M, Swift I, Toomey P, et al. Local staging and
345 assessment of colon cancer with 1.5 Tesla Magnetic Resonance Imaging. *The British Journal of*
346 *Radiology.* 2016:20160257.
- 347 8. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards
348 complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Standards*
349 *for Reporting of Diagnostic Accuracy. Clin Chem.* 2003 Jan;49(1):1-6.
- 350 9. Low RN. MR imaging of the peritoneal spread of malignancy. *Abdom Imaging.* 2007 May-
351 Jun;32(3):267-83.
- 352 10. Low RN. Magnetic resonance imaging in the oncology patient: evaluation of the extrahepatic
353 abdomen. *Semin Ultrasound CT MR.* 2005 Aug;26(4):224-36.

- 354 11. Taylor FG, Swift RI, Blomqvist L, Brown G. A systematic approach to the interpretation of
355 preoperative staging MRI for rectal cancer. *AJR Am J Roentgenol*. 2008 Dec;191(6):1827-35.
- 356 12. Dighe S, Swift I, Brown G. CT staging of colon cancer. *Clin Radiol*. 2008 Dec;63(12):1372-9.
- 357 13. Kim JH, Beets GL, Kim MJ, Kessels AG, Beets-Tan RG. High-resolution MR imaging for nodal
358 staging in rectal cancer: are there any criteria in addition to the size? *Eur J Radiol*. 2004 Oct;52(1):78-
359 83.
- 360 14. Brown G, Richards CJ, Bourne MW, Newcombe RG, Radcliffe AG, Dallimore NS, et al.
361 Morphologic predictors of lymph node status in rectal cancer with use of high-spatial-resolution MR
362 imaging with histopathologic comparison. *Radiology*. 2003;227(2):371-7.
- 363 15. Smith NJ, Barbachano Y, Norman AR, Swift RI, Abulafi AM, Brown G. Prognostic significance
364 of magnetic resonance imaging-detected extramural vascular invasion in rectal cancer. *Br J Surg*.
365 2008 Feb;95(2):229-36.
- 366 16. Dighe S, Blake H, Koh MD, Swift I, Arnaout A, Temple L, et al. Accuracy of multidetector
367 computed tomography in identifying poor prognostic factors in colonic cancer. *Br J Surg*. 2010
368 Sep;97(9):1407-15.
- 369 17. Williams GT, Quirke P, Shepherd NA. Dataset for colorectal cancer (2nd edition): The Royal
370 College of Pathologists 2007.
- 371 18. FOXTROT Collaborative. FOXTROT Protocol version 6.0. [PDF]: Birmingham Clinical Trials Unit;
372 2012 [updated 9/7/2012; cited 2015 15/12/2015]; FOXTROT trial protocol].
- 373 19. Merkel S, Mansmann U, Siassi M, Papadopoulos T, Hohenberger W, Hermanek P. The
374 prognostic inhomogeneity in pT3 rectal carcinomas. *Int J Colorectal Dis*. 2001 Sep;16(5):298-304.

375 20. Andre T, Boni C, Navarro M, Tabernero J, Hickish T, Topham C, et al. Improved Overall
376 Survival With Oxaliplatin, Fluorouracil, and Leucovorin As Adjuvant Treatment in Stage II or III Colon
377 Cancer in the MOSAIC Trial. *Journal of Clinical Oncology*. 2009 Jul 1;27(19):3109-16.

378 21. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or
379 partial credit. *Psychological Bulletin*. 1968;70(4):213-20.

380 22. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. *Journal of*
381 *the American Statistical Association*. 1927;22(158):209-12.

382 23. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement
383 or partial credit. *Psychol Bull*. 1968 Oct;70(4):213-20.

384 24. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R
385 Foundation for Statistical Computing; 2015.

386 25. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research.
387 Evanston, Illinois: Northwestern University; 2013.

388 26. Therneau T. A Package for Survival Analysis in S. 2.38 ed 2015.

389 27. Smith NJ, Bees N, Barbachano Y, Norman AR, Swift RI, Brown G. Preoperative computed
390 tomography staging of nonmetastatic colon cancer predicts outcome: implications for clinical trials.
391 *Br J Cancer*. 2007 Apr 10;96(7):1030-6.

392 28. Burton S, Brown G, Bees N, Norman A, Biedrzycki O, Arnaout A, et al. Accuracy of CT
393 prediction of poor prognostic features in colonic cancer. *Br J Radiol*. 2008 Jan;81(961):10-9.

394 29. Lim M, Hussain Z, Howe A, Storey R, Petty D, Haselden J, et al. The oncological outcome after
395 right hemicolectomy and accuracy of CT Scan as a preoperative tool for staging in right sided colonic
396 cancers. *Colorectal Dis*. 2012 Oct 16.

397 30. Norgaard A, Dam C, Jakobsen A, Ploen J, Lindebjerg J, Rafaelsen SR. Selection of colon cancer
398 patients for neoadjuvant chemotherapy by preoperative CT scan. Scand J Gastroentero. 2014
399 Feb;49(2):202-8.

400 31. Rollven E, Holm T, Glimelius B, Lorinc E, Blomqvist L. Potentials of high resolution magnetic
401 resonance imaging versus computed tomography for preoperative local staging of colon cancer. Acta
402 Radiol. 2013 Apr 2.

403

404

405

Table legends

Table 1: Accuracy, sensitivity and specificity of MRI and CT for identifying T3 or greater disease

Table 2: Accuracy, sensitivity and specificity of MRI and CT for identifying T3c or greater disease

Table 3: Accuracy, sensitivity and specificity of MRI and CT for identifying N1 or greater disease

Table 4: Accuracy, sensitivity and specificity of MRI and CT for identifying EMVI

Table 5: Incidence of metastatic disease by risk factor (whole group)

Table 6: Cox proportional hazards regression models of disease-free survival, by modality and reader, stratified for adjuvant chemotherapy

Figure legends

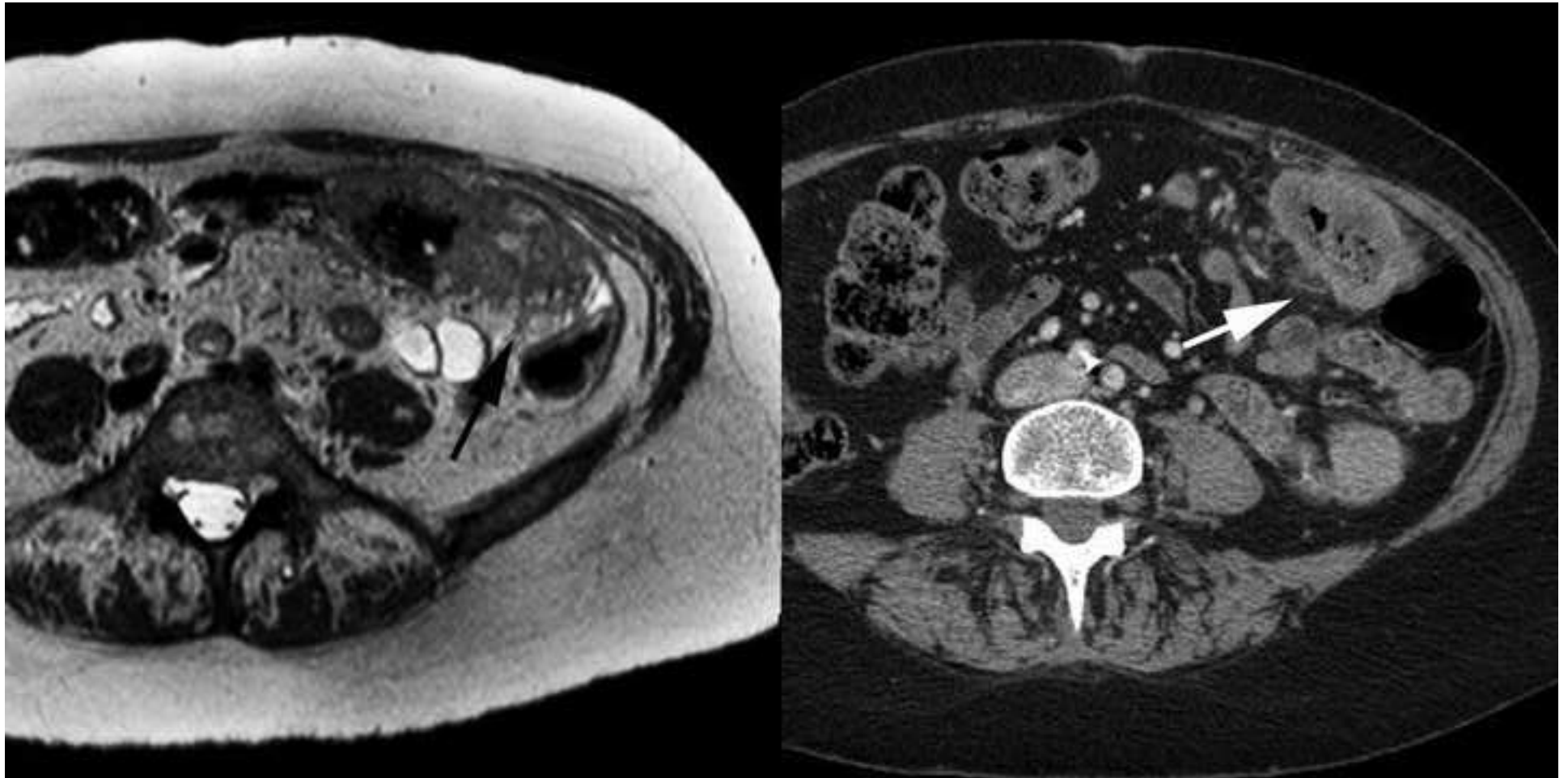
Figure 1. T2 weighted 3T MRI axial (A) and sagittal (B) images, and CT axial (C) and sagittal (D) images of T2 tumour (arrows), confirmed on histopathology.

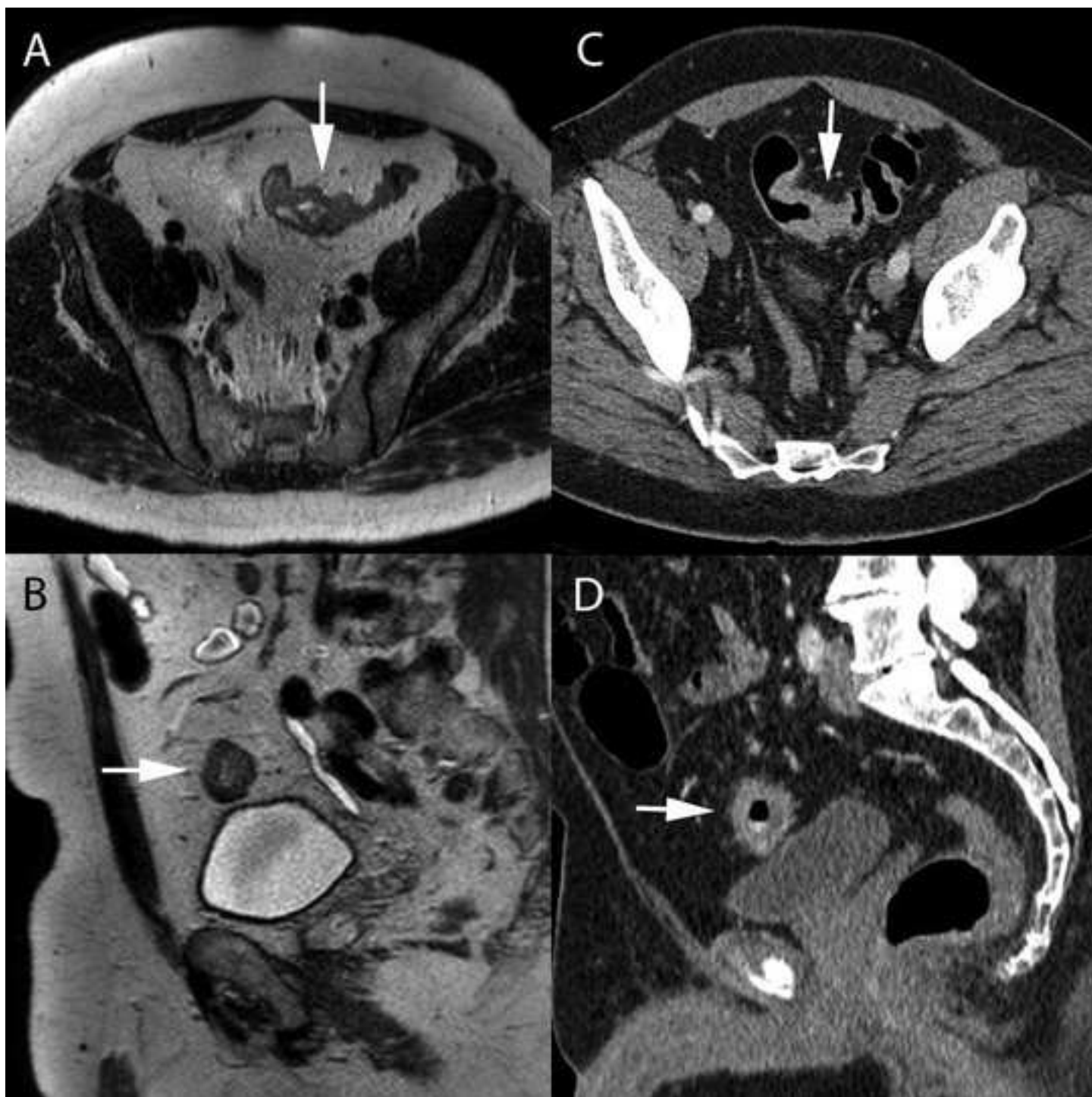
Figure 2. T2 weighted 3T MRI image (left) and CT image (right) of T3 tumour (arrows), confirmed on histopathology.

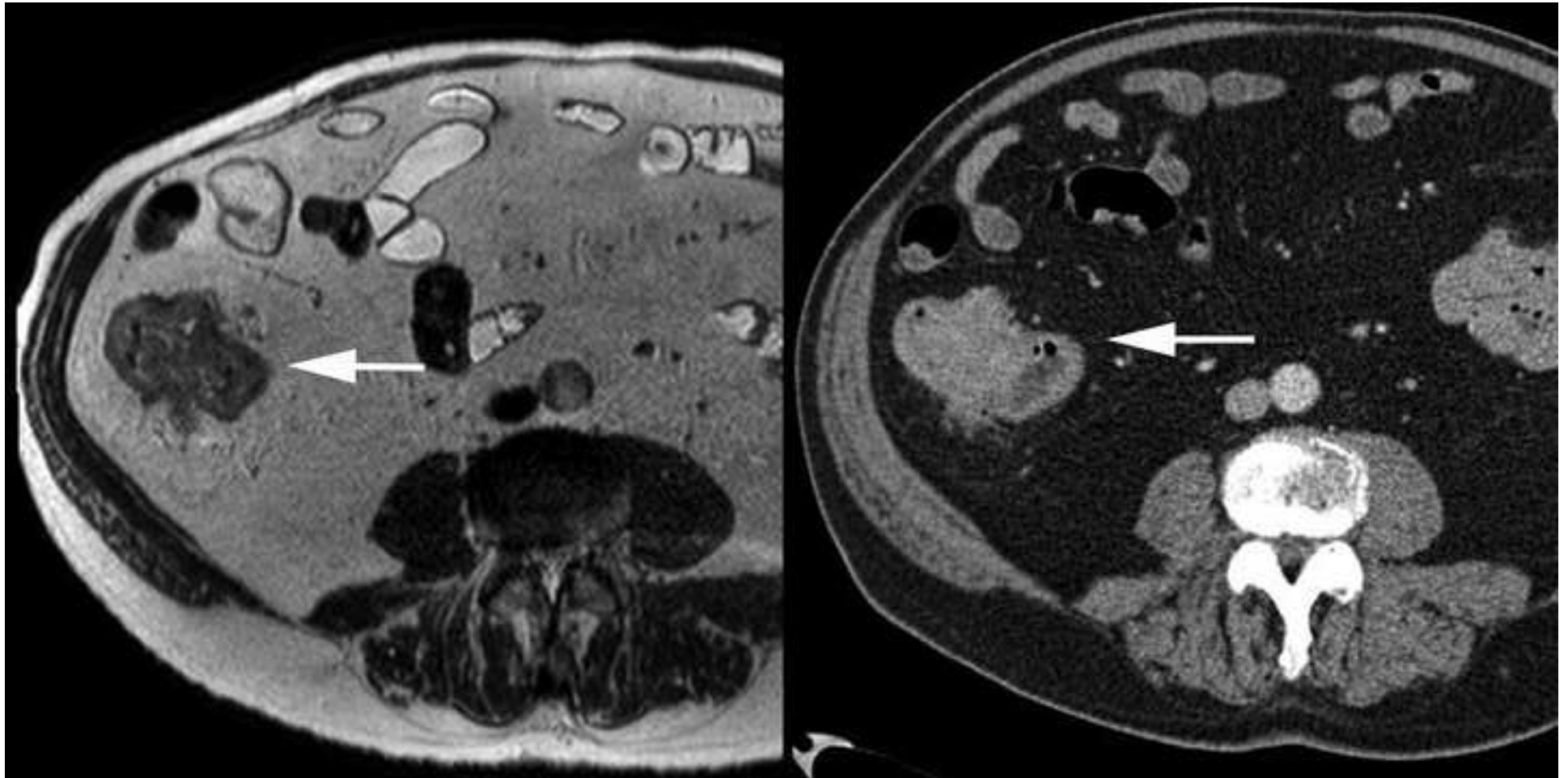
Figure 3. T2 weighted 3T MRI image (left) and CT image (right) of T4 tumour (arrows), confirmed on histopathology.

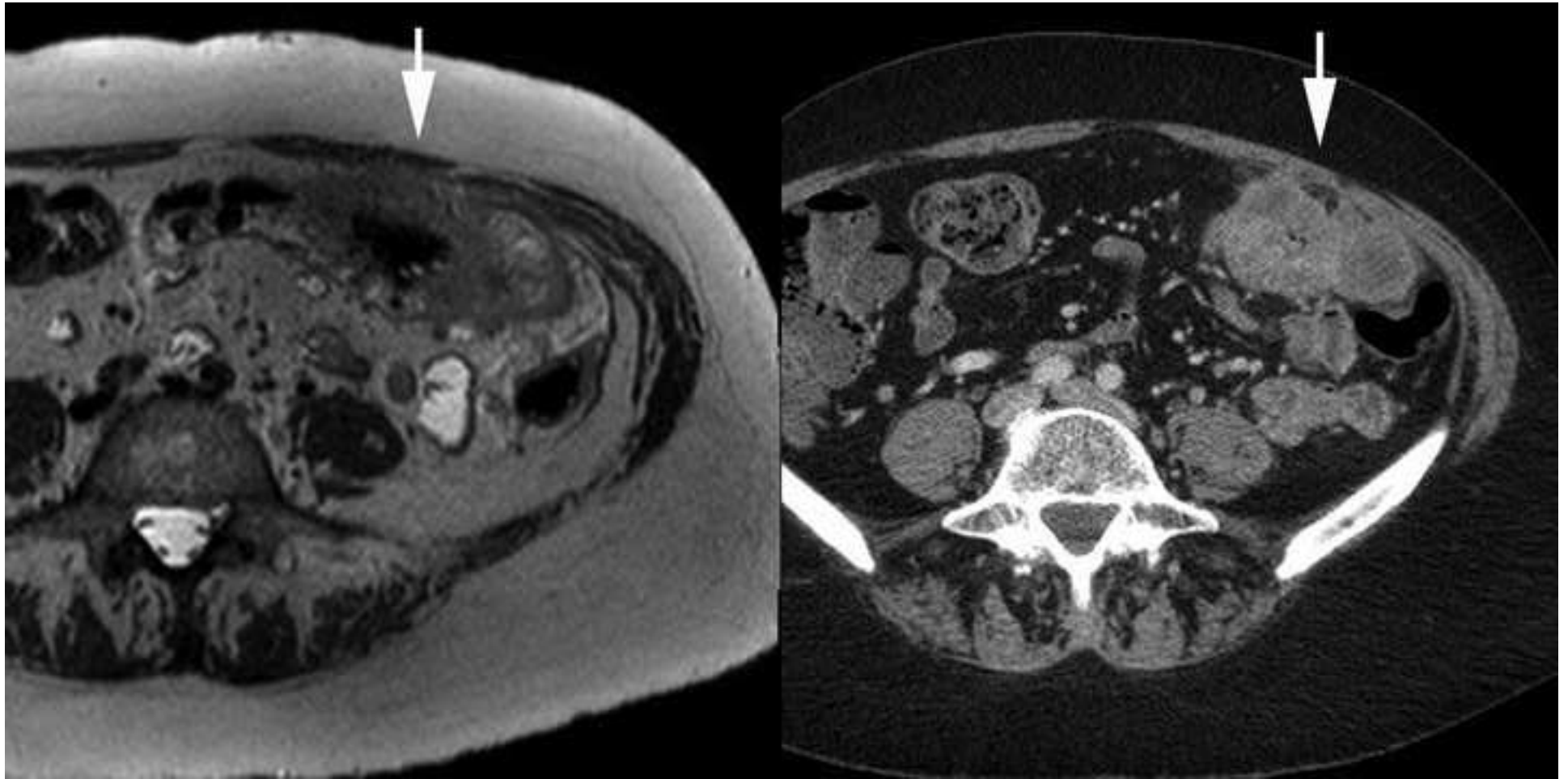
Figure 4. T2 weighted 3T MRI coronal (left) and sagittal (right) images of a metastatic lymph node deposit (arrows), confirmed on histopathology.

Figure 5. T2 weighted 3T MRI image (left) and CT image (right) of extramural vascular invasion (arrows), confirmed on histopathology.









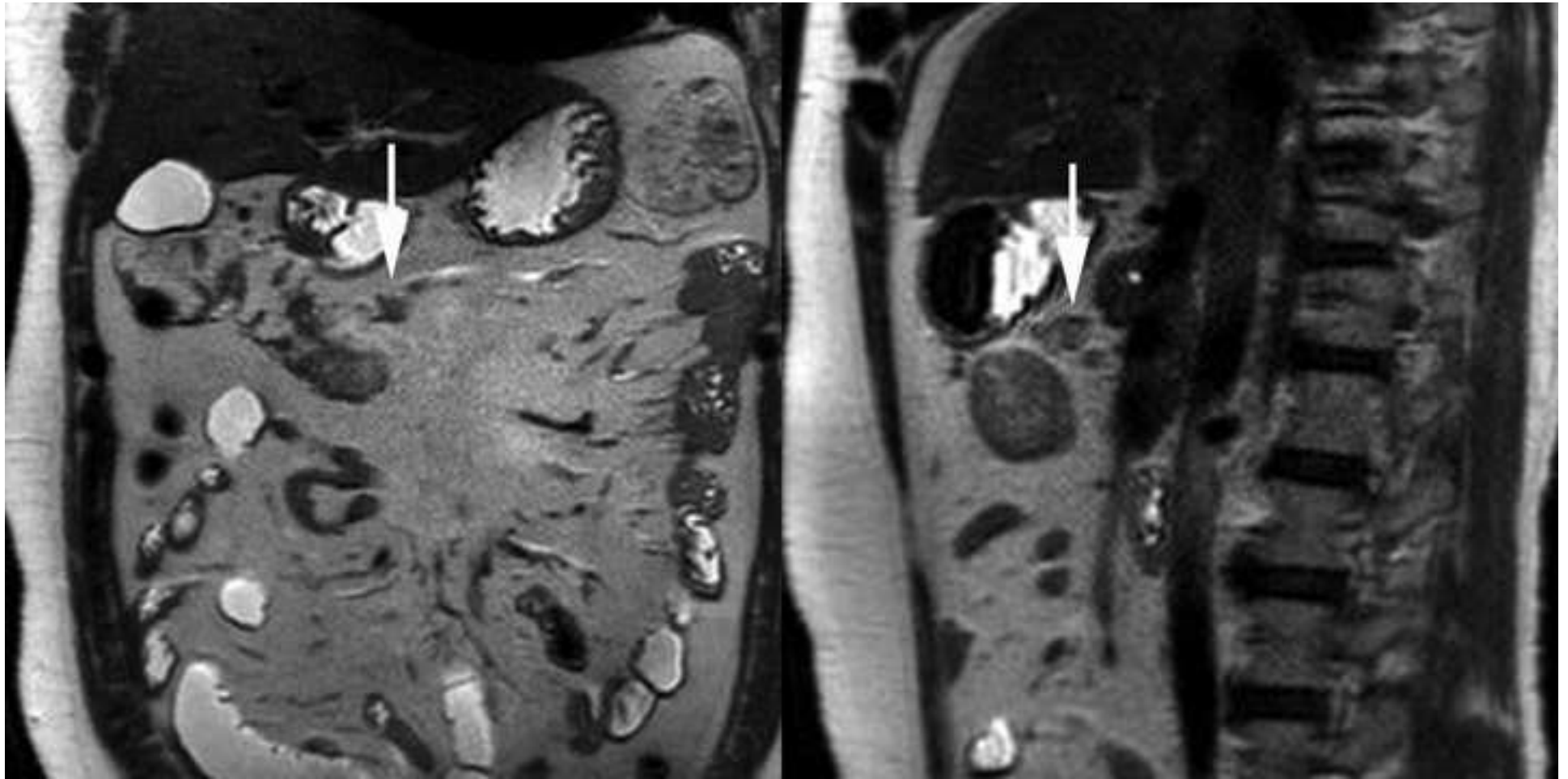


Table1

Reader	Scan	Accuracy % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Reader 1	CT	72 (58-82)	81 (67-90)	36 (15-65)
	MRI	75 (62-85)	79 (64-88)	64 (35-85)
Reader 2	CT	66 (53-77)	69 (54-81)	55 (28-79)
	MRI	57 (43-69)	50 (36-64)	82 (52-95)

Table2

Reader	Scan	Accuracy % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Reader 1	CT	70 (56-80)	70 (48-85)	70 (53-83)
	MRI	75 (62-85)	75 (53-89)	76 (59-87)
Reader 2	CT	77 (64-87)	60 (39-78)	88 (73-95)
	MRI	79 (67-88)	55 (34-74)	94 (80-98)

Table3

Reader	Scan	Accuracy % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Reader 1	CT	62 (48-74)	64 (43-80)	60 (42-75)
	MRI	62 (48-74)	41 (23-61)	77 (59-88)
Reader 2	CT	56 (42-68)	73 (52-87)	43 (27-61)
	MRI	63 (50-75)	45 (27-65)	77 (59-88)

Table4

Reader	Scan	Accuracy % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Reader 1	CT	79 (66-88)	63 (41-81)	88 (73-95)
	MRI	75 (62-85)	74 (51-88)	76 (59-87)
Reader 2	CT	54 (41-67)	47 (27-68)	58 (41-73)
	MRI	75 (62-85)	37 (19-59)	97 (85-99)

Table5

Modality	Risk factor	Odds Ratio (95% CI)	p-value (Fisher's exact test)
	Male	2.1 (0.45-9.4)	0.32
	Chemo	4.0 (0.89-22)	0.048*
	Age >= 70	1.6 (0.36-8.6)	0.53
CT Reader 1	T3 or greater	0.31 (0.060-1.6)	0.12
	T3c or greater	0.88 (0.19-3.9)	1
	N1 or greater	0.65 (0.14-2.9)	0.74
	EMVI positive	1.9 (0.38-8.6)	0.48
CT Reader 2	T3 or greater	1.2 (0.26-6.4)	1
	T3c or greater	3.4 (0.71-16)	0.081
	N1 or greater	1.2 (0.26-6.4)	1
	EMVI positive	0.88 (0.19-3.9)	1
MRI Reader 1	T3 or greater	0.54 (0.12-2.6)	0.49
	T3c or greater	0.97 (0.20-4.3)	1
	N1 or greater	1.2 (0.21-5.4)	1
	EMVI positive	0.97 (0.20-4.3)	1
MRI Reader 2	T3 or greater	2.3 (0.52-11)	0.32
	T3c or greater	3.3 (0.63-17)	0.12
	N1 or greater	2.58 (0.56-12)	0.17
	EMVI positive	2.29 (0.30-15)	0.37
Histopathology	T3 or greater	1.4 (0.24-16)	1
	T3c or greater	8.6 (1.7-59)	0.0044*
	N1 or greater	2.3 (0.52-11)	0.32
	EMVI positive	3.2 (0.71-16)	0.095

Table6

Modality	Risk factor	Hazard ratio (95% CI)	p-value (Wald)
CT Reader 1	T3c or greater	1.3 (0.31-5.5)	0.73
	N1 or greater	0.41 (0.13-1.3)	0.12
	EMVI positive	1.1 (0.27-4.6)	0.89
CT Reader 2	T3c or greater	2.2 (0.64-7.4)	0.21
	N1 or greater	0.98 (0.28-3.4)	0.97
	EMVI positive	0.45 (0.12-1.7)	0.23
MRI Reader 1	T3c or greater	0.76 (0.11-5.3)	0.78
	N1 or greater	2.6 (0.61-11)	0.20
	EMVI positive	0.33 (0.054-2.0)	0.23
MRI Reader 2	T3c or greater	3.6 (0.97-14)	0.056
	N1 or greater	2.6 (0.86-8.2)	0.091
	EMVI positive	0.46 (0.087-2.4)	0.35
Histopathology	T3c or greater	5.4 (1.3-23)	0.024*
	N1 or greater	1.0 (0.31-3.2)	1.0
	EMVI positive	0.58 (0.13-2.6)	0.48

Highlights

- T3c or greater disease is the adverse feature most reliably identified on imaging
- T3c or greater disease on histopathology predicts significantly worse DFS
- Accuracy for identification of adverse features was similar on CT and 3T MRI