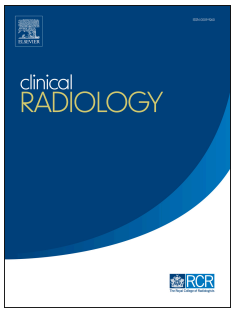
Accepted Manuscript



Inter-observer agreement of radiologists assessing the response of rectal cancers to pre-operative chemoradiation using the MRI tumour regression grading (mrTRG)

M.R.S. Siddiqui, K.L. Gormly, J. Bhoday, S. Balyansikova, N.J. Battersby, M. Chand, S. Rao, P. Tekkis, A.M. Abulafi, G. Brown, Professor

PII: S0009-9260(16)30072-1

DOI: 10.1016/j.crad.2016.05.005

Reference: YCRAD 4361

To appear in: *Clinical Radiology*

Received Date: 14 October 2015

Revised Date: 13 March 2016

Accepted Date: 3 May 2016

Please cite this article as: Siddiqui MRS, Gormly KL, Bhoday J, Balyansikova S, Battersby NJ, Chand M, Rao S, Tekkis P, Abulafi AM, Brown G, Inter-observer agreement of radiologists assessing the response of rectal cancers to pre-operative chemoradiation using the MRI tumour regression grading (mrTRG), *Clinical Radiology* (2016), doi: 10.1016/j.crad.2016.05.005.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

MANUSCRIPT TITLE PAGE

Inter-observer agreement of radiologists assessing the response of rectal cancers to pre-operative chemoradiation using the MRI tumour regression grading (mrTRG).

Author list

1) Muhammed Rafay Sameem Siddiqui[1,2,6]
2) Kirsten L Gormly[4]
3) Jemma Bhoday[1,2,6]
4) Svetlana Balyansikova[2]
5) Nicholas J Battersby[2]
6) Manish Chand[5]
7) Sheela Rao[2]
8) Paris Tekkis[3,6]
9) Al-Mutaz Abulafi[1]
10) Gina Brown[2,6]

Affiliations

[1] Department of Colorectal Surgery, Croydon University Hospital, Croydon, UK, CR7 7YE.
[2]Department of Radiology, Royal Marsden Hospital, Sutton, Surrey, UK, SM2 5PT.
[3]Department of Surgery, Royal Marsden Hospital, Fulham Rd, London, UK, SW3 6JJ.
[4]Dr Jones and Partners, Adelaide, South Australia.
[5]Department of Surgery, University College London, London, UK
[6]Imperial College London, London, UK

Corresponding author:
Professor Gina Brown
Department of Radiology, The Royal NHS Foundation Trust, Downs Road, Sutton, Surrey, SM2 5PT.
Phone: 0208 915 6067
Fax: 0208 915 6721
E-mail:gina.brown@rmh.nhs.uk

Funding:
The corresponding author is funded by a grant from the pelican center

Contributors:
Karen Thomas, Statistician, Royal Marsden Hospital, UK

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Interobserver agreement of radiologists assessing the response of rectal cancers to preoperative chemoradiation using the MRI tumour regression grading (mrTRG)

M. R. S. Siddiqui[1,2,6], K. L. Gormly[4], J. Bhoday[1,2,6], S. Balyansikova[2], N. J. Battersby[2], M. Chand[5], S. Rao[2], P Tekkis[3,6], A.-M. Abulafi[1], G. Brown[2,6]

[1] Department of Colorectal Surgery, Croydon University Hospital, Croydon CR7 7YE, UK

[2]Department of Radiology, Royal Marsden Hospital, Sutton, Surrey SM2 5PT, UK

[3]Department of Surgery, Royal Marsden Hospital, Fulham Rd, London SW3 6JJ, UK

[4]Dr Jones and Partners, Adelaide, South Australia

[5]Department of Surgery, University College London, London, UK

[6]Imperial College London, London, UK

*Guarantor and correspondent: G. Brown, Department of Radiology, The Royal NHS Foundation Trust, Downs Road, Sutton, Surrey, SM2 5PT, UK. Tel.: 0208 915 6067; fax: 0208 915 6721.

E-mail address: gina.brown@rmh.nhs.uk

ABSTRACT

AIM: To investigate whether the magnetic resonance imaging (MRI) tumour regression grading (mrTRG) scale can be taught effectively resulting in a clinically reasonable interobserver agreement (κ>0.4; moderate to near perfect agreement).

MATERIALS AND METHODS: This study examines the interobserver agreement of mrTRG, between 35 radiologists and a central reviewer. Two workshops were

organised for radiologists to assess regression of rectal cancers on MRI staging scans. A range of mrTRGs on 12 patient scans were used for assessment.

RESULTS: Kappa agreement ranged from 0.14–0.82 with a median value of 0.57 (95% CI: 0.37–0.77) indicating good overall agreement. Eight (26%) radiologists had very good/near perfect agreement (κ>0.8). Six (19%) radiologists had good agreement (0.8≥κ>0.6) and a further 12 (39%) had moderate agreement (0.6≥κ>0.4). Five (16%) radiologists had a fair agreement (0.4≥κ>0.2) and two had poor agreement (0.2>κ). There was a tendency towards good agreement (skewness: 0.92). In 65.9% and 90% of cases the radiologists were able to correctly highlight good and poor responders, respectively.

CONCLUSIONS: The assessment of the response of rectal cancers to chemoradiation therapy may be performed effectively using mrTRG. Radiologists can be taught the mrTRG scale. Even with minimal training, good agreement with the central reviewer along with effective differentiation between good and intermediate/poor responders can be achieved. Focus should be on facilitating the identification of good responders. It is predicted that with more intensive interactive cased-based learning a κ>0.8 is likely to be achieved. Testing and retesting is recommended.

INTRODUCTION

The treatment for rectal cancer has improved and has led to better survival outcomes over the last three decades. The reasons for this are multifactorial and include better understanding of pelvic anatomy[1] and surgical techniques[2], earlier diagnoses[3], neoadjuvant or adjuvant therapies[4], and improved imaging[5]. Response

of tumours to neoadjuvant therapy has also allowed more sphincter-sparing procedures[6] with the additional potential for deferral of surgery [7,8].

The degree of tumour response has been shown to be an important prognosticating factor [9,10]. This response may be classified by several methods including: downstaging, most commonly according to the TNM classification[11]; downsizing, usually by Response Evaluation Criteria In Solid Tumours (RECIST)[12]; and, by regression grading[13,14]. Tumour regression appears to be an independent predictor for survival;[1,15] however, there are several scales evident in the literature[13,14,16-24]. This has resulted in confusion as to the precise definition of a “poor”, “intermediate” or “good” responder;[25] consequently, there is a wide variation in the reported disease-free survival (DFS) and overall survival (OS) [16,17,26-41].

More recently magnetic resonance imaging (MRI) has been used for tumour regression grading (mrTRG) with a more representative DFS of 31–68% and OS 27–59% for poor responders and DFS of 64–83% and OS of 72–90% for good responders [42,43]. The mrTRG scale can be accurately taught and utilised by other experienced gastrointestinal radiologists[44], achieving a κ of 0.6 for mrTRG, which was better than the interobserver agreement reported for T-staging at MRI [44]. The purpose of this article is to investigate the interobserver agreement between a central reviewer and 35 radiologists newly taught in mrTRG assessment during a training workshop.

MATERIALS AND METHODS

Hypothesis and sample size

The hypothesis of the present study was that the mrTRG scale can effectively be taught if a clinically reasonable interobserver agreement (κ>0.4; moderate to near

perfect agreement) was achieved and would be rejected if the agreement was $\kappa<0.4$ (poor/fair agreement).

The technique described by Gwet[45] was used to determine the sample size, which assumes a chance probability of 0%. To achieve a minimum κ of 0.4 with a relative error of 20%, a minimum sample size of 156 is required. There were 35 radiologists at two separate workshops, and so a sample of 12 patients was chose, ensuring 420 comparisons made with the central reviewer.

Patients and imaging

A senior radiologist reviewed a patient and imaging database at a cancer centre. Twelve patients with rectal cancer who had undergone long-course chemoradiation therapy (CRT) as part of their cancer treatment protocol between 2008 and 2013 were included. Cases were selected if baseline and post-CRT (at 4-6 weeks) scans were available. The MRI images were examined to ensure a clear and fair representation of the mrTRG scale (Table 1). Anonymised scans were paired as pre- and post-treatment for assessment of levels of agreement.

MRI protocol

A 1.5 T MRI system with phased-array coils was used for all post-CRT scans. T2-weighted large field of view (FOV) axial, and high-resolution small FOV sagittal, oblique axial (perpendicular to the lumen of rectal wall) and oblique coronal sequences were obtained. High-resolution MRI are defined as voxel size <1.5 and sufficient signal-to-noise ratio (SNR) [46]. Pre-treatment scans were performed using a similar protocol.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Reference radiologist standard

The reference standard for comparison was a senior radiologist with 20 years of reporting MRI examinations used to stage rectal cancer. The radiologist developed the mrTRG scale and routinely reported these imaging scans, which were previously performed with histopathological correlation [5]. The same mrTRG scale was used by the central reviewer and the participating radiologists.

Participating radiologists

Thirty-five radiologists from Australia and New Zealand participated in two separate whole-day workshops on rectal cancer staging, including post-CRT assessment. The radiologists ranged from fellows to senior consultants. The participants included those with no experience in reporting rectal MRI to experienced gastrointestinal radiologists. None of the radiologists had prior experience of reporting using the mrTRG criteria. Delegates were not expected to be fully trained in mrTRG on completion, although an additional 30-minute case review session was provided after the assessment, which increased delegates' levels of confidence and understanding. The workshop included a 15–30 minute lecture on the assessment of tumour characteristics after CRT, focusing on assessment according to the mrTRG scale. Following this, the radiologists assessed the pre- and post-treatment scans of 12 patients and gave them an mrTRG score using a standardised proforma. The detail in the proforma used the same mrTRG scale (Fig. 1). The two proformas differed only in the first being more detailed, adding extra points that the radiologist could consider in their assessment of the case. The mrTRG score in both cases relied on the same tumour visible score, which is the information used to give the mrTRG score. There was a slight difference in the TRG score 2 between "minimal" and

"none" on first and second proformas, but the other scores were all unchanged. The additional points of fibrosis, mucin, and lymph nodes on the first proforma were for consideration only. On the second course, these were removed to encourage people to concentrate only on the relevant question of remaining tumour visible. No other pathological or clinical information was given to those participating. Images were assessed on a high-definition reporting monitor. The participants performed the assessment independently and were blinded to the correct assessment until the end of proforma completion. After handing in the proforma, the answers were available and several cases were reviewed in a group setting. No data were collected to identify individual radiologists. Completed data forms were collated and then compared to the reference standard assessment performed by a senior radiologist.

mrTRG scale

The mrTRG scale is based on a regression scale originally described for postoperative resection specimens[13] and uses a five-point scale (Table 1) [42]. Lower TRG scores refer to greater regression and the system further divides the categories into type of response (complete, good, moderate, slight, and none). Using the five-point scale, mrTRG can classify response into good and poor according to survival outcomes [42-44,47]. Good responders are those patients with mrTRG 1 and 2, whereas intermediate/poor responders are mrTRG 3–5 as defined for the purposes of this study.

Data synthesis

Data were tabulated and entered onto a spreadsheet. All statistical analyses were performed using Microsoft Excel version 14.4.5 (2011; Microsoft, Redmond, WA

USA) [48] and IBM SPSS version 22.0 (2013; IBM, NY, USA)[49]. Cohen's κ level was used to calculate the interobserver agreement between the reference standard and individual radiologists and also to obtain an overall value. A value of $p<0.05$ was chosen as the significance level for κ statistics and tests whether the agreement is due to chance, therefore, if $p<0.05$ it proves the null hypothesis that the agreement would not be expected by chance alone[50]. The value of κ statistics was interpreted according to Altman[51]. Agreement lies between 0 and 1, where 0 is indicative of no agreement and 1 indicates complete agreement. "Very good/near perfect" agreement is considered as a κ of 0.81–1.00; "good" agreement as a κ of 0.61–0.80; "moderate" agreement as a κ of 0.41–0.60; "fair" agreement as a κ of 0.21–0.40; "poor" agreement as a κ of <0.2. Rarely, a negative κ is observed, which indicates that the interobserver agreement is less than would be expected by chance and is interpreted as no agreement[52]. A measure of skewness to assess normal distribution was calculated to investigate whether the trend was towards a higher or lower κ and considered significant if the standard error of the skewness was less than half the overall value of the skew[53]. Forest plots were used for graphical display using Comprehensive Meta-Analysis (Biostate, Englewood, NJ, USA)[54].

RESULTS

Thirty-five radiologists completed the proforma for mrTRG scoring (Fig. 1). Two radiologists misunderstood the proforma (achieving a negative κ) and were excluded from the analysis. There were seven missing data entries leaving 389 assessments on 12 patients. There were six male and six female patient images. One patient had mrTRG 1, three had mrTRG 2, three had mrTRG 3, three had mrTRG 4, and two patients had an mrTRG5. This entailed four "good" responders (mrTRG 1–2) and

eight “intermediate/poor” responders (mrTRG 3–5) as defined for the purpose of this study.

Patient demographics

Patient demographics can be seen in Table 2. There were six men and six women. Mean age was 75 years (54–93 years). One patient had metastases at presentation and was a poor responder (mrTRG 4). Five patients were deferred for surgery, four of which were considered good responders (mrTRG 1 and 2). One patient (originally mrTRG 3) was deferred for surgery after a repeat MRI showed ongoing response. All patients who had a good response were given consolidation chemotherapy and underwent close follow-up.

Ability to differentiate between good and intermediate/poor responders

Results of the κ statistic for each radiologist can be seen in Table 3 and Fig. 2. Data from two radiologists were excluded because they misunderstood the proforma (κ = –0.04 and –0.24). Kappa agreement ranged from 0.14–0.82 with a median value of 0.57 (confidence interval [CI]: 0.37–0.77) indicating an overall good agreement (weighted pooled κ was 0.63 as seen in Fig. 2). Eight (8/33; 26%) radiologists had very good/near perfect agreement (κ>0.8). Six (6/33; 19%) radiologists had good agreement (0.8>κ>0.6), and a further 12 (12/33; 39%) had moderate agreement (0.6>κ>0.4). Five (5/33; 16%) radiologists had a fair agreement, and two radiologists had a poor agreement. From the results there was a marked negative skew (–0.92, SE=0.40) indicating that the κ-values tended towards good agreement for mrTRG (Fig. 3).

Variation between workshops

The median interobserver agreement from the first workshop was 0.61 (range=0.38–0.82) and for the second it was 0.53 (range=0.14–0.80). As the comparisons were individual comparisons with the reference standard, it was not statistically feasible to directly compare the two; however, the difference does not constitute a drop in the kappa agreement categories and would be considered within an acceptable variance.

Ability to identify good and poor responders as separate groups

According to central review, there were four patients with mrTRG 1–2 (good responders); with 33 radiologists there were a total of 129 assessments (three assessments were missing). In 65.9% of cases, the radiologists were able to identify good responders correctly in agreement with the study standard. There were eight patients with mrTRG 3–5 (intermediate/poor responders); with 33 radiologists there were a total of 260 assessments (four assessments were missing). In 90% of cases, the radiologists were able to identify intermediate/poor correctly in agreement with the study standard. This suggests that radiologists are better at identifying poor responders than identifying good responders; this has a marked subsequent effect on the interobserver agreement.

DISCUSSION

Identifying an accurate and reproducible assessment of regression after CRT is an important factor in rectal cancer management. Although this has usually been conducted with pathological assessment systems, mrTRG has the advantage of

ACCEPTED MANUSCRIPT

potentially affecting management prior to surgery, and therefore, provides a window of opportunity to act on this information.

Main findings

The present study has shown that after a short period of training, most radiologists were able to differentiate between good and intermediate/poor responders. Overall median κ agreement for differentiating good and intermediate/poor response was 0.57 (0.37, 0.77). The relatively low standard error with narrow CIs indicates that these agreement levels are reproducible and result in clinically acceptable levels. Although 90% of radiologists were able to correctly identify intermediate and poor responders with scores of mrTRG3-5, only 66% correctly identified good responders. This indicates radiologists are more likely to misinterpret fibrosis as residual tumour and greater experience is necessary to have the confidence to report no visible tumour.

Importance of this study

Traditionally complete pathological response is considered to be reflective of better long-term outcomes [15]. There has been extensive work on attempting to classify patients into good and intermediate/poor response to try and personalise treatment options and inform follow-up protocols; however, this has remained challenging due to a range of pTRG scales that assess regression[55] using post-surgical specimens [13,14,16-23]. Although traditionally histopathology was considered the reference standard, a reference standard is typically defined as any technique that predicts outcomes accurately. mrTRG has shown better correlation to survival outcomes in the literature, and therefore, is an important tool in directing treatment, and currently,

would be viewed as the preferred reference standard. Given the role of mrTRG, it would be useful to establish whether it can be taught and utilised effectively by other radiologists. The present study has shown that good clinical κ agreement can be achieved with minimal training. Further focus could concentrate on facilitating the ability to identify good responders.

Appraisal of evidence

The results of the present study are comparable with other studies, including the MERCURY study,[43,44] reporting κ agreement of 0.55–0.65. The present results were not dissimilar to these values, and this highlights that the mrTRG scale can be replicated in a range of settings and may be taught effectively by the techniques of standalone workshops; essentially validating its use by other radiologists in different settings.

Furthermore the present study has shown that mrTRG is consistent and reliable in differentiating between good and intermediate/poor responders. This is higher than reported histological grading systems, which tend toward poor agreement using different histological scales with overall κ values ranging from 0.28–0.38 and approximate median κ values for different scales of 0.24, 0.42, and 0.58 [56]. One issue regarding pathology TRG scales is that there is a perception that it may not actively affect the ongoing management of patients despite its mandatory requirement in reports [57]. The use of mrTRG allows the multidisciplinary team (MDT) to potentially change management decisions preoperatively and consider the use of consolidation chemotherapy and non-operative therapy [58].

Strengths of the study

The main strength of the present study is the sample size of 35 radiologists and 12 patient images giving an effective sample size of 413 (seven missing assessments). This ensured effective statistical analyses to be performed. The teaching setting was part of a workshop attended by radiologists who had an interest in learning or improving their skills in reporting rectal cancer MRI, but none had prior experience in using this mrTRG system. The images were high quality and assessed on reporting monitors, allowing the assessments to be performed according to the mrTRG scale without any confounding image-related factors. The radiologists were also blinded to the reference standard until the end of the assessment. The advantage of establishing mrTRG as the reference standard assessment is to effectively offer management specifically tailored to patients and may include the option of non-operable management or potentially further chemoradiotherapy, with a view to increasing sphincter-saving procedures.

Limitations and heterogeneity of this study

The short period of teaching and assessment may falsely downgrade the κ agreement. Ideally, the assessment could be extended to multiple workshops with an initial period of central review, as there is likely to be a learning curve,[43] even though the initial results are encouraging. As part of a full-day course on rectal MRI staging, the mrTRG was only presented as a brief lecture and there was no hands-on case-based teaching until after the assessments had been completed. The lecture and proforma were also altered slightly between the two workshops. The radiologists had differing degrees of experience, including some who had no prior experience reporting rectal MRI, which may reflect the range of agreement for individual radiologists.

The costs of MRI following treatment may not always be reimbursed in different healthcare systems. Rationing of healthcare spending and limitation of MRI examinations until the test can be proven definitively as a clinical necessity is also a factor in some centres. The lack of confidence in radiological interpretation of post-treatment imaging has also been cited as a contributing factor [59,60].

Implications for clinical practice and future work

The present study indicated that post-treatment assessment of tumour regression (mrTRG) can be taught effectively in a short time period and as the body of evidence increases regarding patient assessment using MRI, subsequent implementation and adoption of the mrTRG scale may be relatively seamless. The use of mrTRG-directed management to offer this stratified approach to treatment will be tested in the multicentre randomised TRIGGER trial (magnetic resonance Tumour Regression Grade (mrTRG) as a novel biomarker to stratify between Good and poor responders following chemoradiotherapy in Rectal cancer). This imminent trial will investigate the initial MRI to guide neoadjuvant therapy requirement and operative planning, including influencing factors such as mrEMVI. A second MRI after neoadjuvant therapy will examine mrTRG. A good responder may be offered deferral of surgery and the poor responders will be offered further treatment or surgery.

In conclusion, the assessment of the response of rectal cancers to chemoradiation therapy may be performed effectively using mrTRG. Radiologists can be taught the mrTRG scale. Even with minimal training, good agreement levels with the central reviewer along with effective differentiation between good and intermediate/poor responders can be achieved. Focus should be on facilitating the identification of

good responders. It is predicted that with more intensive interactive cased-based learning a $\kappa>0.8$ is likely to be achieved. Testing and retesting is recommended.

ACKNOWLEDGEMENTS

The corresponding author was funded by a grant from the Pelican Centre. The authors thank Karen Thomas, Royal Marsden Hospital, UK, for statistical help.

REFERENCES

1. MacGregor TP, Maughan TS, Sharma RA. Pathological grading of regression following neoadjuvant chemoradiation therapy: the clinical need is now. *J Clin Pathol*2012; 65(10): 867-71.

2. Heald RJ, Moran BJ, Ryall RD, Sexton R, MacFarlane JK. Rectal cancer: the Basingstoke experience of total mesorectal excision, 1978–1997. *Arch Surg* 1998; 133(8): 894-9.

3. Logan RF, Patnick J, Nickerson C, et al. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut* 2012; 61(10): 1439-46.

4. Chua YJ, Barbachano Y, Cunningham D, et al. Neoadjuvant capecitabine and oxaliplatin before chemoradiotherapy and total mesorectal excision in MRI-defined poor-risk rectal cancer: a phase 2 trial. *Lancet Oncol* 2010; 11(3): 241-8.

5. Brown G, Richards CJ, Newcombe RG, et al. Rectal carcinoma: thin-section MR imaging for staging in 28 patients. *Radiology* 1999; 211(1): 215-22.

6. De Nardi P, Carvello M. How reliable is current imaging in restaging rectal cancer after neoadjuvant therapy? *W J Gastroenterol* 2013; 19(36): 5964-72.

ACCEPTED MANUSCRIPT

7. Hawkes EA, Cunningham D, Tait D, Brown G, Chau I. Neoadjuvant chemotherapy alone for early-stage rectal cancer: an evolving paradigm? *Semin Radiat Oncol* 2011; 21(3): 196-202.

8. Habr-Gama A, Perez RO, Nadalin W, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg* 2004; 240(4): 711-7; discussion 7-8.

9. Marijnen CA, Glimelius B. The role of radiotherapy in rectal cancer. *Eur J Cancer* 2002; 38(7): 943-52.

10. Pahlman L, Hohenberger W, Gunther K, Fietkau R, Metzger U. Is radiochemotherapy necessary in the treatment of rectal cancer? *Eur J Cancer* 1998; 34(4): 438-48.

11. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* 2010; 17(6): 1471-4.

12. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45(2): 228-47.

13. Mandard AM, Dalibard F, Mandard JC, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 1994; 73(11): 2680-6.

14. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *International journal of colorectal disease* 1997; 12(1): 19-23.

ACCEPTED MANUSCRIPT

15. Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol*2010; 11(9): 835-44.

16. Min BS, Kim NK, Pyo JY, et al. Clinical impact of tumor regression grade after preoperative chemoradiation for locally advanced rectal cancer: subset analyses in lymph node negative patients. *Journal of the Korean Society of Coloproctology* 2011; 27(1): 31-40.

17. Bujko K, Kolodziejczyk M, Nasierowska-Guttmejer A, et al. Tumour regression grading in patients with residual rectal cancer after preoperative chemoradiation. *Radiother Oncol* 2010; 95(3): 298-302.

18. Glynne-Jones R, Anyamene N. Just how useful an endpoint is complete pathological response after neoadjuvant chemoradiation in rectal cancer? *International journal of radiation oncology, biology, physics* 2006; 66(2): 319-20.

19. Junker K, Muller KM, Bosse U, Klinke F, Heinecke A, Thomas M. [Apoptosis and tumor regression in locally advanced non-small cell lung cancer with neoadjuvant therapy]. *Der Pathologe* 2003; 24(3): 214-9.

20. Rodel C, Martus P, Papadoupolos T, et al. Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. *J Clin Oncol.* 2005; 23(34): 8688-96.

21. Schneider PM, Baldus SE, Metzger R, et al. Histomorphologic tumor regression and lymph node metastases determine prognosis following neoadjuvant radiochemotherapy for esophageal cancer: implications for response classification. *Ann Surg* 2005; 242(5): 684-92.

ACCEPTED MANUSCRIPT

22. Wheeler JM, Warren BF, Mortensen NJ, et al. Quantification of histologic regression of rectal cancer after irradiation: a proposal for a modified staging system. *Dis Colon Rectum* 2002; 45(8): 1051-6.

23. Wittekind C, Tannapfel A. [Regression grading of colorectal carcinoma after preoperative radiochemotherapy. An inventory]. *Der Pathologe* 2003; 24(1): 61-5.

24. Japanese Society for Cancer of the Colon and Rectum (JSCCR) (1997) Japanese classification of colorectal carcinoma, 1st English edn. Kanehara & Co, Tokyo.

25. Perez RO. Why do we need another tumor regression grading system for rectal cancer after neoadjuvant therapy? *Dis Colon Rectum* 2015; 58(1): 1-2.

26. Avallone A, Delrio P, Pecori B, et al. Oxaliplatin plus dual inhibition of thymidilate synthase during preoperative pelvic radiotherapy for locally advanced rectal carcinoma: long-term outcome. *International journal of radiation oncology, biology, physics* 2011; 79(3): 670-6.

27. Beddy D, Hyland JM, Winter DC, et al. A simplified tumor regression grade correlates with survival in locally advanced rectal carcinoma treated with neoadjuvant chemoradiotherapy. *Ann Surg Oncol* 2008; 15(12): 3471-7.

28. Eich HT, Stepien A, Zimmermann C, et al. Neoadjuvant radiochemotherapy and surgery for advanced rectal cancer : prognostic significance of tumor regression. *Strahlentherapie und Onkologie : Organ der Deutschen Rontgengesellschaft [et al]* 2011; 187(4): 225-30.

29. Elezkurtaj S, Moser L, Budczies J, et al. Histopathological regression grading matches excellently with local and regional spread after neoadjuvant therapy of rectal cancer. *Pathology, research and practice* 2013; 209(7): 424-8.

30. Gambacorta MA, Valentini V, Morganti AG, et al. Chemoradiation with raltitrexed (Tomudex) in preoperative treatment of stage II-III resectable rectal cancer: a phase II study. *International journal of radiation oncology, biology, physics* 2004; 60(1): 130-8.

31. Giralt J, Tabernero J, Navalpotro B, et al. Pre-operative chemoradiotherapy with UFT and Leucovorin in patients with advanced rectal cancer: a phase II study. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 2008; 89(3): 263-9.

32. Hermanek P, Merkel S, Hohenberger W. Prognosis of rectal carcinoma after multimodal treatment: ypTNM classification and tumor regression grading are essential. *Anticancer research* 2013; 33(2): 559-66.

33. Horisberger K, Hofheinz RD, Palma P, et al. Tumor response to neoadjuvant chemoradiation in rectal cancer: predictor for surgical morbidity? *Int J Colorect Dis* 2008; 23(3): 257-64.

34. Huebner M, Wolff BG, Smyrk TC, Aakre J, Larson DW. Partial pathologic response and nodal status as most significant prognostic factors for advanced rectal cancer treated with preoperative chemoradiotherapy. *W J Surg* 2012; 36(3): 675-83.

35. Lim SB, Yu CS, Hong YS, et al. Failure patterns correlate with the tumor response after preoperative chemoradiotherapy for locally advanced rectal cancer. *J Surg Oncol* 2012; 106(6): 667-73.

36. Pucciarelli S, Toppan P, Friso ML, et al. Complete pathologic response following preoperative chemoradiation therapy for middle to lower rectal cancer is not a prognostic factor for a better outcome. *Dis Colon Rectum* 2004; 47(11): 1798-807.

37. Roy P, Serra S, Kennedy E, Chetty R. The prognostic value of grade of regression and oncocytic change in rectal adenocarcinoma treated with neo-adjuvant chemoradiotherapy. *J Surg Oncol* 2012; 105(2): 130-4.

38. Shin JS, Jalaludin B, Solomon M, Hong A, Lee CS. Histopathological regression grading versus staging of rectal cancer following radiotherapy. *Pathology* 2011; 43(1): 24-30.

39. Suarez J, Vera R, Balen E, et al. Pathologic response assessed by Mandard grade is a better prognostic factor than down staging for disease-free survival after preoperative radiochemotherapy for advanced rectal cancer. *Colorectal Dis.*2008; 10(6): 563-8.

40. Vallbohmer D, Bollschweiler E, Brabender J, et al. Evaluation of histological regression grading systems in the neoadjuvant therapy of rectal cancer: do they have prognostic impact? *Int J Colorect Dis* 2012; 27(10): 1295-301.

41. Winkler J, Zipp L, Knoblich J, Zimmermann F. Simultaneous neoadjuvant radiochemotherapy with capecitabine and oxaliplatin for locally advanced rectal cancer. Treatment outcome outside clinical trials. *Strahlenther Onkol.* 2012; 188(5): 377-82.

42. Shihab OC, Taylor F, Salerno G, et al. MRI predictive factors for long-term outcomes of low rectal tumours. *Ann Surg Oncol* 2011; 18(12): 3278-84.

43. Patel UB, Brown G, Rutten H, et al. Comparison of magnetic resonance imaging and histopathological response to chemoradiotherapy in locally advanced rectal cancer. *Ann Surg Oncol* 2012; 19(9): 2842-52.

44. Patel UB, Taylor F, Blomqvist L, et al. Magnetic resonance imaging-detected tumor response for locally advanced rectal cancer predicts survival outcomes: MERCURY experience. *J Clin Oncol.* 2011; 29(28): 3753-60.

45. Gwet K. Sample size determination in Inter-rater reliability discussion corner. Available at: http://agreestat.com/blog_irr/sample_size_determination.html. Accessed 4 October 2014.

46. Brown G. Rectal Carcimona Staging: a practical approach. RSNA 2010. Available at: http://www.royalmarsden.nhs.uk/consultants-teams-wards/staff/Documents/gina-brown-rectal-carcinoma-staging.pdf. Accessed 30 October 2014.

47. Patel UB, Blomqvist LK, Taylor F, et al. MRI after treatment of locally advanced rectal cancer: how to report tumor response—the MERCURY experience. *AJR Am J Roentgenol* 2012; 199(4): W486-95.

48. Microsoft. *Microsoft Excel*. Redmond, WA: Microsoft, 2003.

49. IBM Corp. Released 2013. *IBM SPSS Statistics for Mac*, version 22.0. Armonk, NY: IBM Corp.

50. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005; 37(5): 360-3.

51. Altman DG. *Practical statistics for medical research.* London: Chapman & Hall, 1990.

52. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012; 22(3): 276-82.

53. Tabachnick BG, Fidell LS. *Using multivariate statistics*, 3rd edn. New York: Harper Collins, 1996.

54. Borenstein M, Hedges L, Higgins J, Rothstein H. *Comprehensive Meta-analysis*, version 2. Englewood NJ: Biostat, 2005.

ACCEPTED MANUSCRIPT

55. Siddiqui M, Bhoday J, Abulafi AM, Tekkis P, Brown G. Defining poor response after neoadjuvant therapy in patients with rectal cancer. *Br J Surg* 2015; 102(Suppl. 1): 9-118.

56. Chetty R, Gill P, Govender D, et al. International study group on rectal cancer regression grading: interobserver variability with commonly used regression grading systems. *Hum Pathol* 2012; 43(11): 1917-23.

57. Valentini V, Aristei C, Glimelius B, et al. Multidisciplinary Rectal Cancer Management: 2nd European Rectal Cancer Consensus Conference (EURECA-CC2). *Radiother Oncology* 2009; 92(2): 148-63.

58. O'Neill BD, Brown G, Heald RJ, Cunningham D, Tait DM. Non-operative treatment after neoadjuvant chemoradiotherapy for rectal cancer. *Lancet Oncol* 2007; 8(7): 625-33.

59. Chen CC, Lee RC, Lin JK, Wang LW, Yang SH. How accurate is magnetic resonance imaging in restaging rectal cancer in patients receiving preoperative combined chemoradiotherapy? *Dis Colon Rectum* 2005; 48(4): 722-8.

60. Kuo LJ, Chern MC, Tsou MH, et al. Interpretation of magnetic resonance imaging for locally advanced rectal carcinoma after preoperative chemoradiation therapy. *Dis Colon Rectum* 2005; 48(1): 23-8.

Table 1 – Summary of magnetic resonance imaging (MRI) tumour regression scale (mrTRG).

TRG scale	mrTRG (low no.=more regression)
1	No/minimal fibrosis visible (tiny linear scar) and no tumour signal

ACCEPTED MANUSCRIPT

2	Dense fibrotic scar (low signal intensity) but no macroscopic tumour signal *(indicates no or microscopic tumour)*
3	Fibrosis predominates but obvious measureable areas of tumour signal visible
4	Tumour signal predominates with little/minimal fibrosis
5	Tumour signal only: no fibrosis, includes progression of tumour

Table 2. Patient demographics

Study ID	mrTRG	Sex	Age	Date of diagnosis	Metastases presentation	Deferred for good response	Further treatment (chemo)	Operated
1	3	F	72	01/11/11	No	No	No	Yes
2	4	M	86	01/03/11	Yes	No	Yes	Yes
3	3	F	93	14/04/11	No	Yes	No	No
4	5	F	70	01/11/12	No	No	Yes	No
5	1	F	69	08/11/11	No	Yes	Yes	No
6	2	F	67	28/05/10	No	Yes	Yes	Yes
7	4	M	73	01/06/12	No	No	No	No
8	4	F	88	29/05/13	No	No	No	Yes
9	5	M	81	12/07/12	No	No	Yes	No
10	2	M	81	18/04/11	No	Yes	Yes	No
11	3	M	54	10/05/11	No	No	No	Yes
12	2	M	67	26/08/08	No	Yes	Yes	Yes

ACCEPTED MANUSCRIPT

Table 3. Interobserver agreement for radiologists' ability to identify good and poor responders

Radiologist	Kappa	S.E	Confidence interval	*p*-Value
1	0.82	0.17	0.48, 1.16	<0.001
2	0.38	0.31	–0.24, 0.99	0.24
3	0.80	0.19	0.42, 1.18	<0.01
4	0.53	0.30	–0.06, 1.12	0.11
5	0.80	0.19	0.42, 1.18	<0.01
6	0.82	0.17	0.48, 1.16	<0.001
7	0.61	0.25	0.10, 1.11	<0.05
8	0.53	0.30	–0.06, 1.12	0.11
9	0.63	0.24	0.14, 1.11	<0.05
10	-0.04	0.34	–0.72, 0.65	1.00
11	0.61	0.25	0.10, 1.11	<0.05
12	0.40	0.30	–0.2, 1.00	0.21
13	0.40	0.30	–0.2, 1.00	0.21
14	0.79	0.20	0.40, 1.19	<0.01
15	0.53	0.24	0.07, 1.00	<0.05
16	0.63	0.24	0.16, 1.10	<0.05
17	0.53	0.24	0.07, 1.00	<0.05
18	0.31	0.35	–0.37, 0.99	0.39
19	0.80	0.19	0.43, 1.17	<0.01
20	0.57	0.28	0.03, 1.11	0.06
21	0.80	0.19	0.43, 1.17	<0.01

ACCEPTED MANUSCRIPT

22	0.25	0.31	–0.35, 0.85	0.43
23	0.57	0.28	0.03, 1.11	0.06
24	0.56	0.28	–0.003 :1.12	0.08
25	0.63	0.24	0.16, 1.10	<0.05
26	0.80	0.19	0.43, 1.17	<0.01
27	0.25	0.31	–0.35, 0.85	0.43
28	0.14	0.35	–0.55, 0.83	0.69
29	0.40	0.30	–0.19, 0.99	0.21
30	0.57	0.28	0.03, 1.11	0.06
31	0.31	0.35	–0.37, 0.99	0.39
32	-0.24	0.30	–0.83, 0.35	0.43
33	0.14	0.35	–0.55, 0.83	0.69
34	0.80	0.19	0.43, 1.17	<0.01
35	0.40	0.30	–0.19, 0.99	0.21

ACCEPTED MANUSCRIPT

Figure 1 – Proforma used to assess radiologists understanding of mrTRG

1st group

mrTRG assessment

Tumour visible	None	Minimal	<50%	>50%	Same
Fibrosis	>75%	50-75%	Minimal	None	
Mucin	>50%	Minimal	None		
Lymph Nodes	Smaller	New Mucin	Same		
TRG STAGE:	**1**	**2**	**3**	**4**	**5**

2nd group

Of the **remaining** rectal wall abnormality; assess the balance between tumour and fibrosis

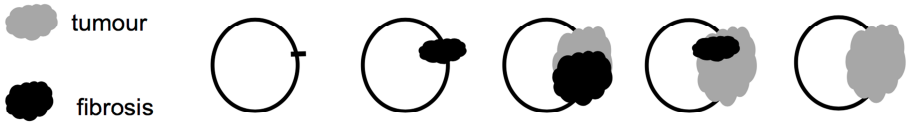
TRG1: no/minimal fibrosis visible (tiny linear scar) and no tumour signal

TRG2: dense fibrotic scar (low signal) but no tumour signal *(indicates no or minimal tumour)*

TRG3: fibrosis predominates but obvious measureable areas of tumour signal visible

TRG4: tumour signal predominates with little / minimal fibrosis

TRG5: tumour signal only – no fibrosis, includes progression of tumour



Tumour visible:	none	none	<50%	>50%	same / worse
TRG:	**1**	**2**	**3**	**4**	**5**

ACCEPTED MANUSCRIPT

Figure 2 – Inter-observer agreement when differentiating between good and poor responders

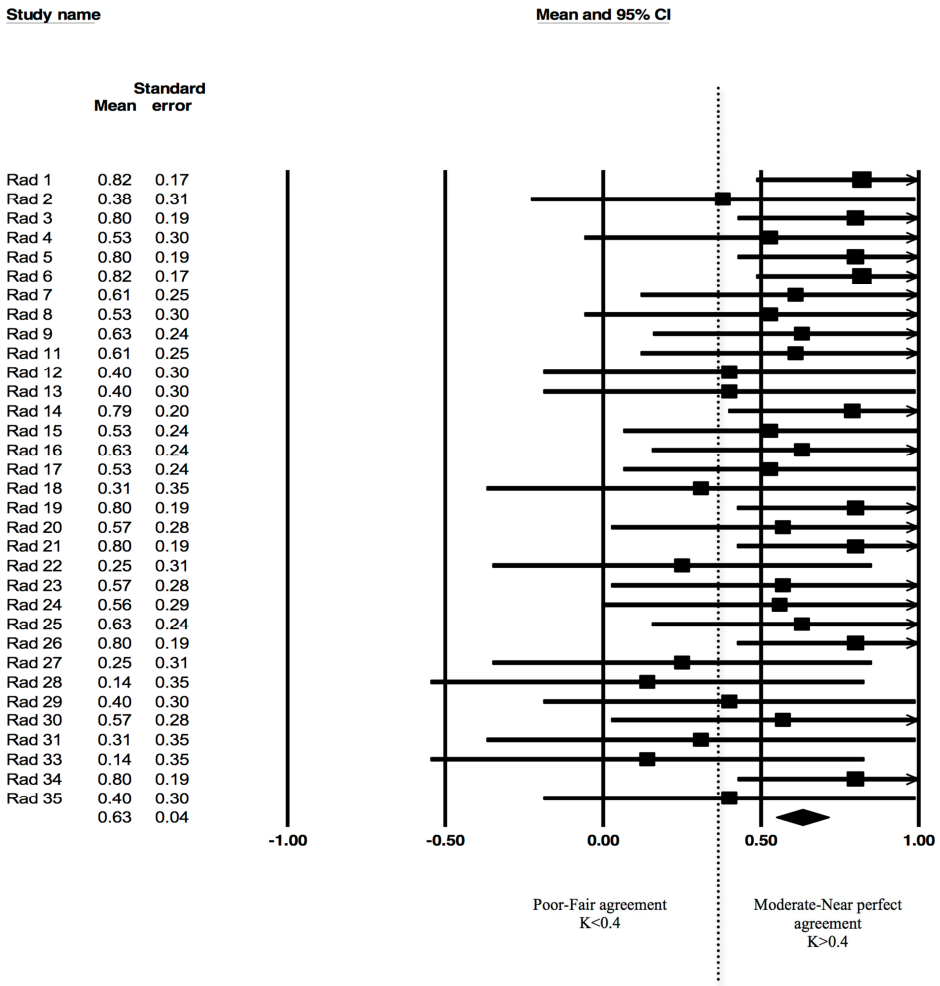
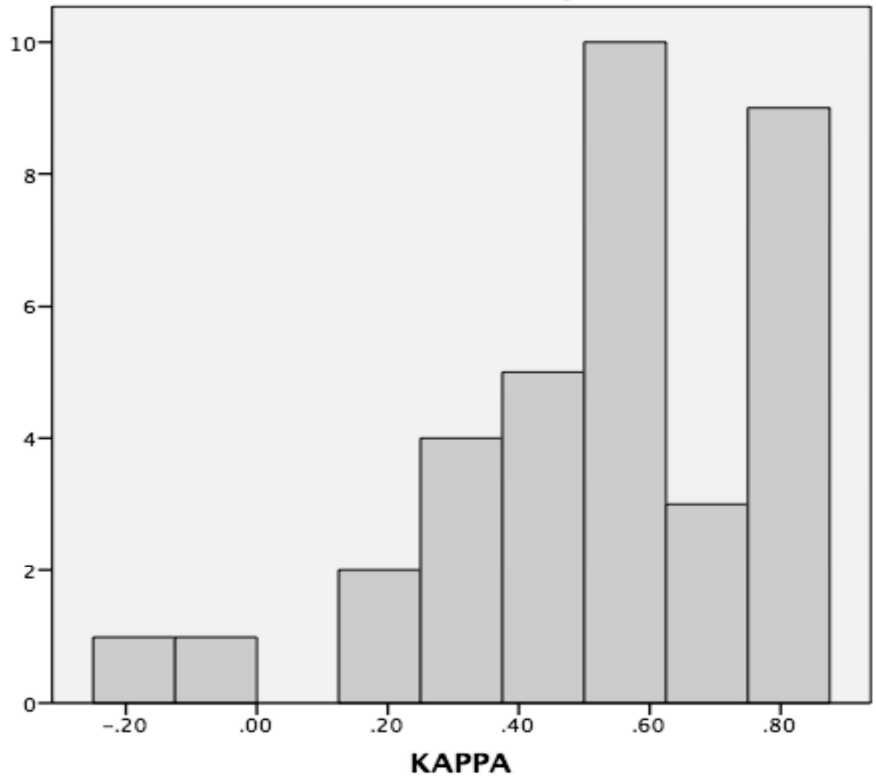


Figure 3 – Distribution of κ values amongst radiologists



ACCEPTED MANUSCRIPT

Highlights:

- Inter-observer agreement of radiologists was assessed, when using the MRI tumour regression scale to determine response of rectal cancers to chemoradiotherapy
- Kappa agreement had a median value of 0.57 (95% CI: 0.37-0.77) indicating an overall good agreement.
- In 65.9% and 90% of cases the radiologists were able to correctly highlight good and poor responders respectively.
- Radiologists can be taught the mrTRG scale and even with minimal training good agreement and effective differentiation between good and intermediate/poor responders can be achieved.
- Focus should be on facilitating the identification of good responders.

ACCEPTED MANUSCRIPT