Radiomics for MRI prediction of tumor response after chemoradiotherapy in rectal cancer

Stuart A Taylor MD MB BS MRCP FRCR

UCL Centre for Medical Imaging - CMI

Division of Medicine

2nd Floor Charles Bell House

43 - 45 Foley Street

London W1W 7TY

Corresponding author

Stuart A Taylor

UCL Centre for Medical Imaging - CMI

Division of Medicine

2nd Floor Charles Bell House

43 - 45 Foley street

London W1W 7TY

Stuart.taylor@ucl.ac.uk

+44 208 138 7949

@profstuartaylor

Word count:1413

Number of references: 10

Patients with locally advanced rectal cancer undergo neoadjuvant chemoradiotherapy to improve local disease control, surgical resectability, and, ultimately, prognosis. There is increasing evidence that around 20% of patients achieve a pathologic complete response (pCR) after chemoradiation alone. Achieving pCR means that no remaining viable tumor is identified during histological analysis of the resection specimen(1). Long-term outcome data is being collated, but a sizable proportion of these patients are essentially cured and can be spared the morbidity and mortality of surgery(1, 2). To avoid unnecessary surgery, such patients must be identified upfront after completing chemoradiotherapy. Once identified, appropriately counseled patients then enter intensive follow-up regimes (often referred to as "wait and watch" programs). These regimes help ensure that any tumor recurrence (which occurs in approximately 25% of patients(1)) can be identified at an early stage and appropriate rescue treatment instigated.

Pelvic MRI, alongside clinical and endoscopic assessment, is crucial in identifying patients who have potentially achieved pCR after chemoradiotherapy and is routinely performed as part of clinical care pathways(3). Response assessment is based on the morphology and signal characteristics of any residual tumor. Currently, no standardized assessment criteria are universally adopted, but the absence of intermediate tumor signal on T2 weighted MRI and/or normalization of the rectal wall with or without low signal fibrosis is the cornerstone of evaluation(3). The MRI tumor regression grade(4) for example uses T2 weighted imaging to grade regression from 1 to 5, with grades 1 (and 2) often interpreted as potential pCR. Adding diffusion-weighted imaging (DWI) to conventional MRI sequences may improve diagnostic performance. The absence of a signal on high b values images (typically > 800sec/mm2) is taken to indicate potential pCR when combined with concurrent evaluation of tumor morphology on T2 weighted MRI(5). A recent meta-analysis reports the sensitivity and specificity of MRI for pCR as 49% (95% CI 33 to 65) and 86% (74 to 93), respectively, using T2 weighted sequences and 62% (43 to 77%) and 89% (80% to 94%) when combining DWI and T2 weighted MRI sequences (6). There was however marked heterogeneity in the data and it is clear more research is needed, particularly around the optimal combination of T2 and DWI criteria.

Although MRI is a powerful tool to appropriately triage patients to surveillance programs, it does have limitations both in missing pCR, resulting in potentially unnecessary surgery, and

over-diagnosing pCR, with the risk of tumor recurrence. These limitations underlie the motivation for researchers investigating whether radiomics techniques can detect patterns in the data beyond those appreciated by the radiological eye, ultimately improving response classification.

In this issue of *Radiology*, Shin and colleagues(7) present a radiomics model for predicting pCR on posttreatment MRI and report promising results. Using a retrospectively identified cohort of 898 patients with locally advanced rectal cancer undergoing MRI after chemoradiotherapy and before surgical resection, the authors developed and evaluated radiomics models based on T2 weighted MRI alone, apparent diffusion coefficient (ADC) MRI maps alone, and a merged model based on both sequences. They employed temporal validation whereby the cohort was split into 592 training sets and 306 test sets according to the date of scan acquisition and used surgical histopathological analysis of the resection specimen as the reference standard. Radiomics software extracted 1132 candidate features for T2 and ADC calculated using MRI. The authors retained only those features with good interobserver reproducibility and reduced redundancy by collapsing highly correlated features into one representative feature. The least absolute shrinkage and selection operator (LASSO) method was used to select the most powerful predictive features from the training set. In the test dataset, the best performing models for predicting pCR were those based on T2 weighted images (sensitivity 80% [95% CI: 71, 89], specificity 68% [95% CI: 62, 76], and the merged model (sensitivity 76% [95% CI: 66, 86] and specificity 71% [95% CI: 68, 77]).

The study has several strengths and adds to the current level of knowledge. The study design included a comparison with radiologist response grading, which many radiomics studies omit. The comparative pooled radiologist sensitivity for prediction of pCR (based on the tumor regression grade of 1 or 2, supplemented by DWI) was 51 % (95% CI: 44, 57), with 89% (95% CI: 87, 92) specificity, consistent with the current literature. Both the T2 and merged models achieved significantly greater sensitivity than radiologists, although specificity for both models was significantly lower.

MRI data was acquired on scans from two different manufacturers (Philips vs Siemens) and at both 1.5T and 3T, which aids generalizability. Using 40 randomly selected lesions, the authors also report good reducibility between two radiologists using software-assisted

region of interest segmentation for radiomic feature extraction. Their method required subjective assessment to exclude “equivocal normal rectal wall” and “mucosal edema”. It is interesting to speculate whether this additional subjective step aided or detracted from the performance of radiomics models and if it will impact on generalizability. High reproducibility is crucial in any radiomics model and key considerations are if and how radiologists should input to region selection, and their interaction, if any, with model outputs. Of note, the segmentation did not include lymph nodes, which is a potential limitation as residual nodal disease is important in clinical decision-making.

The work raises important issues that must be considered in all radiomics research(8). The authors essentially propose a predictive marker for a short-term clinical event, i.e., post chemoradiotherapy pCR using posttreatment MRI. This seems a sensible and clinically meaningful approach and aligns with current staging and treatment pathways. An important limitation however is that non-imaging predictive markers such as endoscopy are not included. As noted above, assessment of pCR is based not just on MRI but also on clinical and endoscopic assessment, and future work must incorporate all relevant clinical predictors. This is crucial to adoption by the clinical community; clinicians know from day to day practice how simple clinical metrics can predict patient outcomes and are unlikely to be persuaded by models that have not considered them(8). Also, as a single-center study, the generalizability to other health care settings remains as yet unknown. The authors, however, derived optimal model cutoff values for their models (in this specific dataset), which can now be tested in larger multicenter studies.

The size of the study is notable and picks up from where similar smaller studies have left off(9, 10). However, the issue of study power must be considered. When powering prognostic markers studies the “rule of 10” is commonly applied i.e. 10 event rates (in this case pCR) are required per predictive variable tested(8). There were 114 “events” in the 592 patient development dataset and 75 in the 306 patient evaluation dataset. Like most radiomics studies, there was no formal power calculation and datasets were collected from the database of one single tertiary institution. There were originally 1132 features obtained from each T2 weighted image and ADC map, although the authors did reduce the variables in their models based on reproducibility and hierarchical feature clustering. The best

performing T2 and merged models eventually included 19 and 27 radiomic features, respectively, so the study risks being potentially underpowered.

The authors used temporal validation (splitting data into training and test sets based on the date of scan acquisition) to develop and evaluate their models. Such an approach is superior to commonly used internal validation (where training and test sets use the same data). But the approach falls short of true external validation in which models are tested using new data from a wide range of institutions and patient samples representative of those in whom the tool will ultimately be employed.

In summary, the prediction of pCR is an important goal in the modern management of rectal cancer. Alongside endoscopy, MRI is a powerful tool in triaging tumor response after chemoradiation although it is limited by relatively modest sensitivity and specificity. In their paper, Shin and colleagues suggest that radiomic models have the potential to improve diagnostic performance although there is a long way to go before such models can be considered for clinical use. Prediction models must include all relevant clinical variables, not just those based on imaging. They must be shown to be reproducible and generalizable. This requires appropriately powered multicenter studies encompassing the variety of data encountered in clinical practice. Assessing the impact on clinical decision-making and, ultimately, patient outcomes is crucial, alongside evaluation of cost-effectiveness. Finally, it is of note that the specificity of the models proposed by Shin et al was significantly less than that of radiologists; models should not be static and need to be updated and improved as new data is accumulated.

References

1. van der Valk MJM, Hilling DE, Bastiaannet E, Meershoek-Klein Kranenbarg E, Beets GL, Figueiredo NL, et al. Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. Lancet (London, England). 2018;391(10139):2537-45. Epub 2018/07/07. doi: 10.1016/s0140-6736(18)31078-x. PubMed PMID: 29976470.

2. Yu G, Lu W, Jiao Z, Qiao J, Ma S, Liu X. A meta-analysis of the watch-and-wait strategy versus total mesorectal excision for rectal cancer exhibiting complete clinical response after neoadjuvant chemoradiotherapy. World journal of surgical oncology. 2021;19(1):305. Epub 2021/10/20. doi: 10.1186/s12957-021-02415-y. PubMed PMID: 34663341; PubMed Central PMCID: PMCPMC8522111.

3. Santiago I, Rodrigues B, Barata M, Figueiredo N, Fernandez L, Galzerano A, et al. Re-staging and follow-up of rectal cancer patients with MR imaging when "Watch-and-Wait" is an option: a practical guide. Insights into imaging. 2021;12(1):114. Epub 2021/08/11. doi: 10.1186/s13244-021-01055-w. PubMed PMID: 34373961; PubMed Central PMCID: PMCPMC8353037.

4. Patel UB, Brown G, Rutten H, West N, Sebag-Montefiore D, Glynne-Jones R, et al. Comparison of magnetic resonance imaging and histopathological response to chemoradiotherapy in locally advanced rectal cancer. Annals of surgical oncology. 2012;19(9):2842-52. Epub 2012/04/25. doi: 10.1245/s10434-012-2309-3. PubMed PMID: 22526897.

5. Lambregts DMJ, Delli Pizzi A, Lahaye MJ, van Griethuysen JJM, Maas M, Beets GL, et al. A Pattern-Based Approach Combining Tumor Morphology on MRI With Distinct Signal Patterns on Diffusion-Weighted Imaging to Assess Response of Rectal Tumors After Chemoradiotherapy. Diseases of the colon and rectum. 2018;61(3):328-37. Epub 2018/01/26. doi: 10.1097/dcr.0000000000000915. PubMed PMID: 29369900.

6. Park SH, Cho SH, Choi SH, Jang JK, Kim MJ, Kim SH, et al. MRI Assessment of Complete Response to Preoperative Chemoradiation Therapy for Rectal Cancer: 2020 Guide for Practice from the Korean Society of Abdominal Radiology. Korean journal of radiology.

2020;21(7):812-28. Epub 2020/06/12. doi: 10.3348/kjr.2020.0483. PubMed PMID: 32524782; PubMed Central PMCID: PMCPMC7289703.

7. Shin J, Seo N, Baek SE, Son NH, Lim JS, Kim NK, Koom WS, Kim S. Large-Scale MRI Radiomics Model Validation to Predict Pathological Complete Response After Chemoradiotherapy in Rectal Cancer. Radiology. (in press).

8. Halligan S, Menu Y, Mallett S. Why did European Radiology reject my radiomic biomarker paper? How to correctly evaluate imaging biomarkers in a clinical setting. European radiology. 2021. Epub 2021/05/19. doi: 10.1007/s00330-021-07971-1. PubMed PMID: 34003349.

9. Horvat N, Veeraraghavan H, Khan M, Blazic I, Zheng J, Capanu M, et al. MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response after Neoadjuvant Therapy. Radiology. 2018;287(3):833-43. Epub 2018/03/08. doi: 10.1148/radiol.2018172300. PubMed PMID: 29514017; PubMed Central PMCID: PMCPMC5978457.

10. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2017;23(23):7253-62. Epub 2017/09/25. doi: 10.1158/1078-0432.Ccr-17-1038. PubMed PMID: 28939744.

Bio

Dr Stuart Taylor is Professor of Medical Imaging at the Centre for Medical Imaging, University College London. His research interests focus on abdominal and pelvic imaging in oncological and inflammatory conditions, particularly using MRI. Professor Taylor is a fellow of the European Society of Gastrointestinal and Abdominal Radiology and sits on their executive committee. He is also a NIHR senior investigator.