Radiotherapy and Oncology 210 (2025) 111032

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Predicting acute diarrhoea in rectal cancer chemoradiotherapy: Secondary analysis of the phase III ARISTOTLE trial

Ying Zhang [a,*], Douglas Brand [a,b], Zhuoyan Shen [a], Mikael Simard [a], Sumeet Hindocha [a,b], Olivia Chohan [a,b], Andre Lopes [c], Rubina Begum [c], Nicholas West [d], Ane Appelt [d,g], Alexandra Gilbert [d], Elizabeth Miles [e], Tim Maughan [f], David Sebag-Montefiore [d], Maria A. Hawkins [a,b,1], Charles-Antoine Collins Fekete [a,1]

[a] Department of Medical Physics and Biomedical Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom
[b] University College London Hospitals NHS Foundation Trust, Radiotherapy Physics, 250 Euston Road, London NW1 2PG, United Kingdom
[c] University College London, Cancer Institute, London, United Kingdom
[d] Leeds Institute of Medical Research, University of Leeds, Leeds, United Kingdom
[e] Mount Vernon Cancer Centre, The Radiotherapy Trials Quality Assurance Group, London, United Kingdom
[f] Department of Molecular and Clinical Cancer Medicine, University of Liverpool, United Kingdom
[g] Department of Medical Physics, Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom

ARTICLE INFO

Keywords:
Rectal cancer
Radiotherapy
Acute diarrhoea
Predictive model
Clinical application
Machine learning

ABSTRACT

Background: Neoadjuvant chemoradiotherapy is a standard treatment for locally advanced rectal cancer, but acute diarrhoea remains a significant side effect, affecting the completion of chemoradiotherapy treatment.
Purpose: This study aimed to predict acute diarrhoea after neoadjuvant chemoradiotherapy for rectal cancer and further develop a strategic tool to individualise rectal cancer treatment.
Materials and methods: The ARISTOTLE trial is a phase III trial comparing capecitabine chemo-radiotherapy (CRT) versus capecitabine-irinotecan CRT as a pre-operative treatment for locally advanced rectal cancer. We included 589 trial patients across 73 institutions. The volume of the AI-segmented small bowel receiving at least 10 Gy (V_{10Gy}) was used alongside the treatment arm, patient age, and performance status in a logistic regression model to predict a more than 2-grade increase in acute diarrhoea toxicity from baseline ($\Delta G \geq 2$). Finally, based on the prediction, we identified a sub-cohort of patients for whom a viable dose decrease would result in a reduction of toxicity, and conversely, we also identified individuals for whom adding irinotecan may not cause toxicity.
Results: The average mean receiver operating characteristic curve (AUROC) for predicting $\Delta G \geq 2$ is 0.71 [95 % CI 0.58–0.82] on the independent test dataset. Based on the prediction, we identified 71 patients (14 %) who could potentially benefit from irinotecan addition without a dose decrease to maintain $\Delta G < 2$, and 77 patients (15 %) who could potentially benefit from irinotecan addition but need a dose decrease to maintain $\Delta G < 2$.
Conclusion: The multi-institutional cohort of 73 centres strengthens the reliability of these findings, demonstrating the model's potential as a strategic tool to individualise rectal cancer treatment while mitigating severe diarrhoea.

Introduction

Acute diarrhoea is a common side effect of chemoradiotherapy (CRT) for rectal cancer, occurring in 12–39 % of patients undergoing preoperative treatment [1–3]. It is caused by intestinal mucosal dysfunction, which leads to impairment of water absorption. Treatment-induced inflammation may cause hyper-mobility and further impairment of intestinal function [4]. Failure to respond to medical management and dietary measures may lead to an inability to complete chemoradiotherapy treatment and/or be associated with chronic diarrhoea.

* Corresponding author.
E-mail address: ying.zhang.18@ucl.ac.uk (Y. Zhang).
[1] Senior authors contributed equally to this work.

https://doi.org/10.1016/j.radonc.2025.111032
Received 16 April 2025; Received in revised form 5 June 2025; Accepted 2 July 2025
Available online 5 July 2025

The establishment of intestinal radiation dose metrics correlating with the risk of acute diarrhoea could direct further efforts to reduce the incidence of this morbid side effect.

The impact of radiation dose on the small bowel and its predictive power on acute diarrhoea has been explored in several studies [1,5–7] using different patient cohorts and small bowel V_{xGy} (volume that receives a radiation dose of x Gy or more). These studies emphasised the significance of a low radiation dose to a high volume of the small bowel as a predictor for acute diarrhoea in rectal cancer treatment. However, the largest patient cohort among these studies consisted of only 203 individuals [7]. This relatively small sample size raises concerns about the generalizability and robustness of the findings. To establish more reliable and comprehensive guidelines for treatment planning, further research involving larger patient cohorts and external validations is required.

This study aims to predict acute diarrhoea grade increases from the baseline scores to grades 2 and above ($\Delta G \geq 2$) using a high-quality clinical trial dataset. We used novel approaches to optimise this analysis including an AI-segmented small bowel contouring to reduce inter-operator variability and using treatment scenario simulation to identify per-patient radiotherapy and concurrent chemotherapy toxicity risk. To the best of our knowledge, this has not been incorporated previously in prediction modelling. These results may contribute to patient selection efforts, allowing clinicians to tailor treatment plans based on individual patient characteristics and predicted toxicity profiles.

Methods and materials

Patient cohort/imaging

This is an institutionally approved post-hoc analysis using data from a phase III trial comparing standard versus novel CRT as the pre-operative treatment for locally advanced rectal cancer (ARISTOTLE, trial identification number: ISRCTN09351447) [8]. This multicentre (73 institutions), open-label trial, enrolled 589 patients between October 2011 and July 2018, randomly assigning (1:1) patients to pre-operative 3D conformal radiotherapy (RT) 45 Gy/25 fractions, combined with either arm A (standard arm): capecitabine 900 mg/m^2 orally twice daily on days of radiotherapy for five weeks or arm B (experimental arm): irinotecan 60 mg/m^2 intravenously once weekly from week 1 to 4 only with capecitabine 650 mg/m^2 orally twice daily on days of radiotherapy for five weeks. The patient characteristics are listed in Table 1.

Of the 589 patients enrolled, 567 had acute diarrhoea toxicity grade recorded based on Common Terminology Criteria for Adverse Events (CTCAE) Version 4.02. A further 67 were excluded due to the following reasons: 62 missing radiotherapy data, 4 received less than 50 % of the planned dose radiotherapy dose (< 20 Gy), and 1 toxicity recording was discontinued prematurely at week 2, indicating non-completion of the treatment regimen. 500 patients recruited from 73 randomising sites remained for the toxicity study.

All patients had assessments for diarrhoea recorded pre-randomisation, pre-treatment, weekly during CRT (week 1–5), at week 6 (1 week after CRT), and week 10. The baseline acute toxicity was pre-treatment diarrhoea.

Auto-segmentation for small bowel

The ARISTOTLE trial was conducted prior to the routine implementation of intensity-modulated radiotherapy (IMRT) and therefore routine small bowel contouring was not mandated in the trial protocol. We required dose-volume metrics of the small bowel for predictive models. We therefore deployed an existing AI-based auto segmentation tool, TotalSegmentator [9] to obtain the necessary small bowel contours. The performance of TotalSegmentator on small bowel is shown in Appendix A.

Table 1
Patient and tumour characteristics. The percentage number represents the fraction of patients within that category that possess a particular feature.

Variables	Arm A (n = 250)		Arm B (n = 250)	
	Median	IQR	Median	IQR
Age (years)	61	53–67	60	53–68
Small bowel V_{10Gy} (cc)	120.2	25.6–220.0	123.1	23.4–221.5
	Number	Percentage	Number	Percentage
Acute diarrhoea toxicity $\Delta G \geq 2$	31	12.4 %	94	37.6 %
Sex				
Female	78	31.2 %	155	62.0 %
Male	172	68.8 %	95	38.0 %
T stage				
T2	17	6.8 %	14	5.6 %
T3	194	77.6 %	187	74.8 %
T4	37	14.8 %	44	17.6 %
Unknown	2	0.8 %	5	2.0 %
N stage				
N0	72	28.8 %	44	17.6 %
N1	106	42.4 %	124	49.6 %
N2	70	28.0 %	75	30.0 %
N3	0	0.0 %	1	0.4 %
Unknown	2	0.8 %	6	2.4 %
Performance status				
1	177	70.8 %	171	68.4 %
0	47	18.8 %	54	21.6 %
Unknown	26	10.4 %	25	10.0 %

Abbreviations: IQR: Interquartile range, T /N stage: MRI-defined tumour/node stage based on 5th AJCC cancer staging manual. Performance status: Eastern Cooperative Oncology Group (ECOG) performance status. 1 means patients are fully active. 0 means patients are ambulatory.
cc: cubic centimetre.

Model evaluation procedure and evaluation metrics

The worst acute toxicity from weeks 1 to 10 was used to calculate the treatment-caused grade increase from baseline, categorised dichotomously between cases where $\Delta G \geq 2$ and cases where $\Delta G < 2$, which was the prediction target in our study. Logistic regression was selected for $\Delta G \geq 2$ prediction due to its simplicity and easy interpretation. The features used include age, ECOG performance, small bowel V10 and arm.

We used a validation strategy (workflow illustrated in Appendix B.3) to assess the performance of logistic regression in predicting $\Delta G \geq 2$. The 500 patients were divided randomly into two groups: an independent test dataset and an internal training dataset. First, institutions were randomly selected to form an independent dataset, ensuring that the patients from these institutions account for at least 10 % of the total patient cohort. The remaining patients formed the internal dataset. The internal dataset was used in 2000 bootstrapping. Each bootstrap sample was generated by randomly selecting patients from the internal dataset with replacement until the sample size matched that of the original internal dataset [10]. The model was trained on each bootstrap sample and was tested on the independent dataset. This whole procedure was repeated 1000 times with different independent datasets to assess logistic regression's predictive performance across various patient populations, enabling a full evaluation of each feature's influence on the probability of $\Delta G \geq 2$, referred to as $P(\Delta G \geq 2)$. Model performance was evaluated using AUROC, recall, specificity, and precision [11] for bootstrapping samples and independent datasets. Results are summarized as mean ± 95 % confidence interval for AUROC, recall, specificity, and precision. SHAP analysis [12] was adopted for model interpretation.

The details related to feature selection, handling data missingness, data curation and outlier detection can also be found in Appendix B.

Model-guided personalised treatment to reduce acute diarrhoea toxicity risk and rationally select for systemic treatment intensification

We deployed the models to investigate two clinical scenarios:

1. Toxicity reduction:
(i) For a given patient receiving rectal CRT with a high predicted toxicity risk, what reduction in the small bowel V_{10Gy} would reduce the $P(\Delta G \geq 2) < 0.5$?
(ii) For patients from arm B (experimental arm) in the 1(i) scenario where dose reduction is not feasible, does the removal of irinotecan make $P(\Delta G \geq 2) < 0.5$?
2. Identifying patients from the standard arm (A) who could safely receive additional irinotecan, without a toxicity penalty:
(i) For a given patient from arm A with a low predicted toxicity risk, does the model predict that irinotecan could be added without increasing the $P(\Delta G \geq 2) > 0.5$?
(ii) For patients in the 2(i) scenario whose $P(\Delta G \geq 2) > 0.5$ after the addition of irinotecan, is there a feasible dose reduction in small bowel V_{10Gy} that would reduce $P(\Delta G \geq 2) < 0.5$?

We used a per-patient analysis to evaluate these hypotheses. Given the variability in model predictions when using different training datasets, relying on a single model for clinical decision-making at this stage poses limitations. To predict outcomes for an individual patient, we employed a bootstrapping approach to generate a distribution of predictions from models trained on different subsets of the patient cohort. The patient of interest was initially excluded from the cohort, and the remaining patients' data were used for 5000 bootstrapping. We generated a distribution of predicted probabilities for $\Delta G \geq 2$ for each held-out patient under the four clinical scenarios by training multiple models on bootstrap datasets. A successful proposed alteration of the treatment would satisfy two criteria: the average $P(\Delta G \geq 2)$ is reduced below 0.5, as well as a statistically different distribution than the non-altered treatment, as calculated from the Baumgartner-Weiss-Schindler (BWS) test ($p < 0.05$) [13]. The Baumgartner-Weiss-Schindler test was used because BWS is more robust to outliers and non-normal data.

Results

Model evaluation

The logistic regression showed consistent performance across training (AUROC: 0.72, 95 % CI: 0.67–0.77) and independent datasets (AUROC: 0.71, 95 % CI: 0.58–0.82) on average. The average recall and specificity of the test sets are 0.66 and 0.68, respectively. The low average precision (~0.40) indicated a high false-positive rate, primarily in irinotecan-treated patients from arm B. Results are presented in Table 2.

SHAP analysis and clinical interpretation

We adopted the SHAP analysis on 5000 fitted models (randomly selected from 2,000,000 models in Section 2.3) to explain the contribution of each feature to the model's output. Fig. 1(a) illustrates the relationship between each feature's value and its corresponding SHAP value (representing the feature's direct influence on the model prediction) for 5000 models trained on different subsets of the cohort. The plots illustrate the uncertainty of a feature' influence on $P(\Delta G \geq 2)$, indicated by the blue shade, while the red line represents the average influence for a given feature across the 5000 models.

Fig. 1(b) summarises the average relationship between features and their SHAP values in a beeswarm plot, enabling reliable clinical interpretation of the overall trends in feature contributions to model predictions. The plot displays features in descending order of importance. The treatment arm emerges as the strongest predictor, in keeping with the expected adverse event profile of irinotecan. Better ECOG performance status (negative SHAP values) correlates with a lower risk of $\Delta G \geq 2$, likely due to better bowel movement and water absorption associated with increased physical activity. Age contributes, with older patients exhibiting higher SHAP values, indicating a greater risk of toxicity. Lower small bowel V_{10Gy} values correspond to lower SHAP values, suggesting that reducing small bowel radiation dose can mitigate the risk of $\Delta G \geq 2$.

Table 2
Performance metrics for predicting Δ Grade ≥ 2 of small bowel toxicity.

AUROC	Precision	Recall	Specificity
Train			
0.72(0.67–0.77)	0.39(0.33–0.45)	0.74(0.67–0.81)	0.61(0.56–0.67)
Test			
0.71(0.58–0.82)	0.40(0.26–0.55)	0.66(0.43–0.86)	0.68(0.54–0.81)

For all metrics: Mean (95% confidence interval) across 2,000,000 models.

Model-guided personalised treatment aiming to minimise acute diarrhoea toxicity risk

We present model-guided personalized treatment for four patients, two for each clinical scenario defined in section 2.4. This individual patient analysis is followed by an investigation of cohort-level statistics to gain a broader understanding of the treatment effects and potential for toxicity reduction across the entire study population.

Fig. 2(a) presents two patients assigned to arm B who experienced $\Delta G \geq 2$. The boxplot in Fig. 2(a) suggests that reducing patient A's V_{10Gy} by 52 % (from 296 cc to 141 cc) could potentially shift the model's prediction from $\Delta G \geq 2$ to $\Delta G < 2$. In contrast, for patient B, the model predicts a high $P(\Delta G \geq 2)$ even with a hypothetical V_{10Gy} of 0 cc, indicating that reducing V_{10Gy} alone may be insufficient to manage toxicity risk for this patient. However, the removal of irinotecan from the treatment plan could allow for $\Delta G < 2$ in patient B.

Fig. 2(b) presents two patients assigned to arm A who exhibited $\Delta G < 2$. We explored the possibility of benefiting from the addition of irinotecan (arm B) by changing their arm label to arm B and investigating the effect of reducing the small bowel V_{10Gy} to maintain $\Delta G < 2$. The boxplot in Fig. 2(b) reveals that when patient C's V_{10Gy} is reduced below 130 cc (25 %), switching the treatment strategy to include irinotecan is not predicted to change the toxicity outcome label. In contrast, for patient D, the model predicts a high $P(\Delta G \geq 2)$, with the hypothetical treatment strategy switch to Arm B, even with a V_{10Gy} of 0 cc. This suggests that reducing V_{10Gy} alone may be insufficient to manage the toxicity risk for patient D due to the influence of other contributing factors.

We applied model-guided personalized treatment for all 500 patients. Fig. 3 illustrates the proportion of patients experiencing changes in the predicted $P(\Delta G \geq 2)$ based on two potential interventions: reducing the small bowel V_{10Gy} and/or modifying the treatment strategy by adding/removing irinotecan.

Table 3 summarizes the predictive features for two groups of patients based on the predicted probabilities from bootstrapped models: 1) cohort of patients who can potentially benefit from the addition of irinotecan without an increase in bowel toxicity and 2) patients for whom using irinotecan would lead to $\Delta G \geq 2$, regardless of the radiation adaptation strategy. The patients who can potentially benefit from irinotecan are predominantly performance status 1, younger, and have lower small bowel V_{10Gy} values.

Among the 500 patients, we identified 71 patients (14.2 %) who could potentially receive additional without the need for bowel radiation dose decrease to maintain $P(\Delta G \geq 2) < 0.5$. 77 patients (15.4 %) can potentially change their predicted toxicity label from 1 to 0 by reducing the small bowel V_{10Gy}. Of these 77 patients, 55 (11 %) require a V_{10Gy}

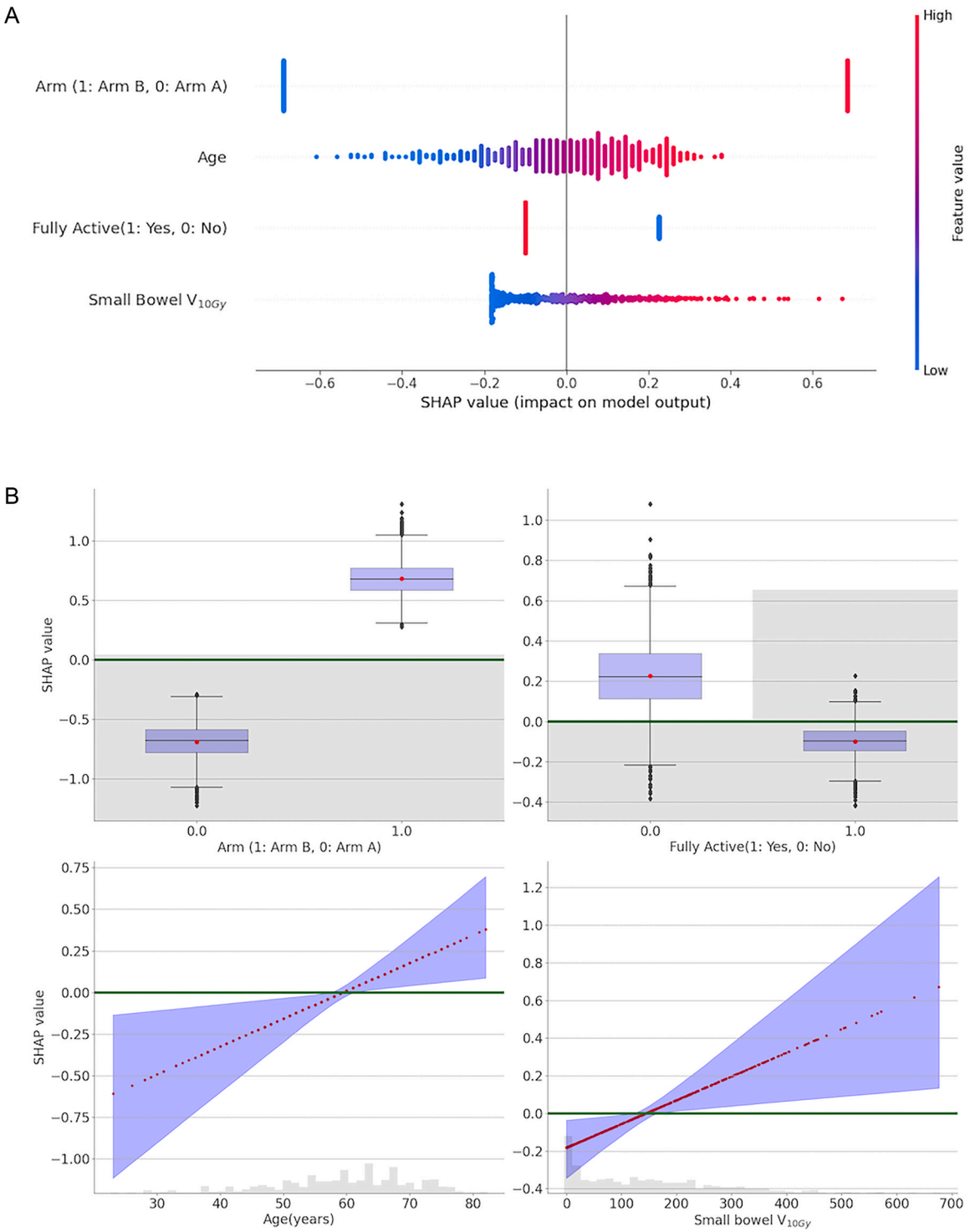

Fig. 1. Model interpretation based on SHAP values. **A)** The dependence plots illustrate the relationship between a feature's value and its corresponding SHAP value, based on the results from 5000 models trained on various datasets. The inset histograms (grey areas) just above the x-axis display the distribution of raw feature values. The mean SHAP values (red dots) and its 95 % CI (blue shade) of a feature for each patient are shown here versus raw feature values. The green horizontal line in each plot represents a SHAP value of 0, indicating no influence on the model's predictions relative to the average prediction. The further a SHAP value lies from zero, the more strongly that feature influences the model's prediction: positive values push the predicted risk upward, whereas negative values pull it downward. **B)** The beeswarm plot visualises the SHAP values for each feature, with individual dots representing patients from the dataset. The colour gradient encodes the raw feature values, with red dots indicating relatively higher values and blue dots representing relatively lower values. The horizontal position of a dot along the x-axis signifies the magnitude and direction of the feature's influence on the model's output. Dots to the right of the origin (x = 0) have positive SHAP values, indicating features that increase the predicted risk of the outcome, while dots to the left of the origin have negative SHAP values, corresponding to features that decrease the predicted risk. The vertical stacking of dots with identical SHAP values enables the visualisation of density and distribution within the dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

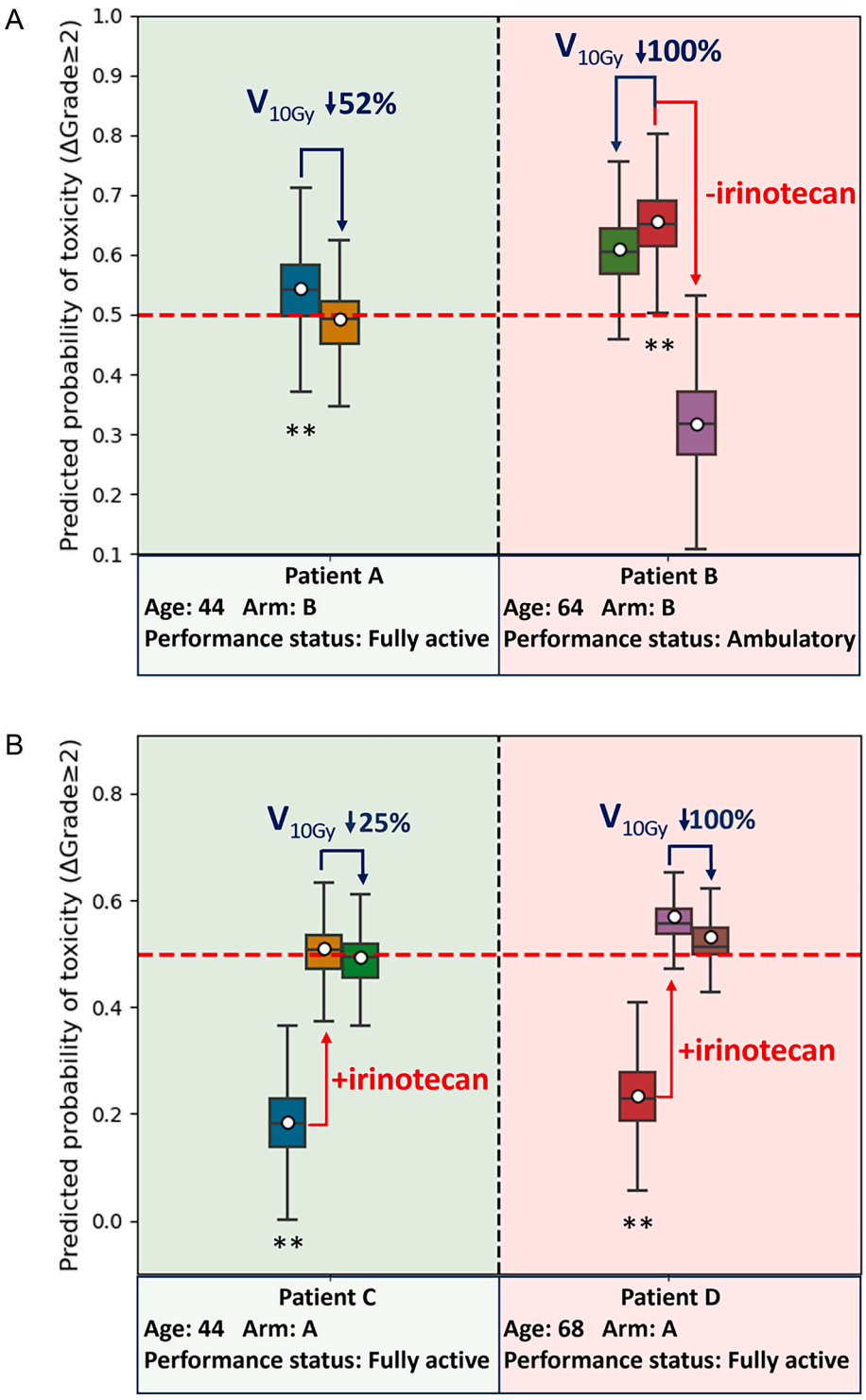

Fig. 2. Two clinical scenarios for model-guided application. **A)** The boxplot illustrates the distribution of predicted probabilities for patients assigned to arm B, considering both their original V_{10Gy} values, reduced V_{10Gy} scenarios and removal of irinotecan. The red line at 0.5 represents the classification threshold separating the two prediction classes. The patient highlighted in green exemplifies a case where a reduction in V_{10Gy} successfully lowers the predicted $P(\Delta G \geq 2)$ to below 50 %. In contrast, the patient marked in red cannot achieve $\Delta G < 2$ by reducing V_{10Gy} alone, even when reduced to zero. However, the removal of irinotecan from the treatment plan would allow this patient to achieve $\Delta G < 2$. **B)** The boxplot shows the distribution of predicted probabilities of experiencing $\Delta G \geq 2$ for the patients who were assigned to arm A, under three scenarios: 1) their original arm and their original small bowel V_{10Gy}, 2) a hypothetical switch to arm B (including irinotecan) and their original V_{10Gy}, and 3) a hypothetical switch to arm B with the reduced V_{10Gy} value. The red line at 0.5 serves as the classification threshold that separates the two prediction classes. The patient highlighted in green is identified by our model as potentially benefiting from a switched treatment strategy (from arm A to arm B) with a reduction in V_{10Gy} to maintain $\Delta G < 2$ while potentially enhancing tumour control through the addition of irinotecan. Conversely, the patient marked in red cannot achieve $\Delta G < 2$ by reducing V_{10Gy} with a switched treatment strategy, according to the model's predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

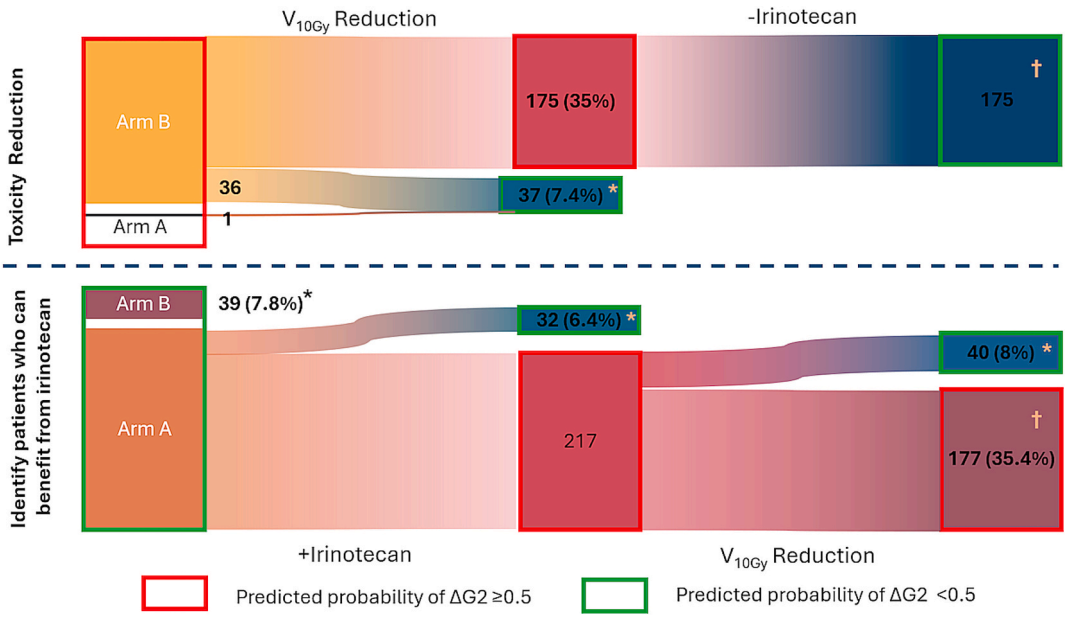

Fig. 3. Sankey diagram to show the proportion of predicted probability changes corresponding to small bowel V_{10Gy} reduction and/or treatment strategy changes. The red boxes indicate the cases whose predicted $P(\Delta G \geq 2) \geq 0.5$ and green boxes indicate the cases whose predicted $P(\Delta G \geq 2) < 0.5$. Patients who can potentially benefit from Irinotecan with $\Delta G < 2$ are marked by *. Patients who use irinotecan would lead to $\Delta G \geq 2$ are marked by †. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Comparison of features for patients who can benefit from *Irinotecan with $\Delta G < 2$ and who cannot.*

	With irinotecan < 0.5 (*)	With irinotecan ≥ 0.5(†)	Statistics (p value)
Proportion of fully active patient in the sub-cohort (%)	95.9	58.8	χ^2 square test: <0.001
	Median (IQR)		
Age(years)	51(42–55)	64(60–69.25)	*t*-test: <0.001
Small bowel V_{10Gy} (cc) (with modified value)	47.2(11.9–90.6)	125.2 (24.1–222.9)	*t*-test: <0.001
Original small bowel V_{10Gy} (cc)	116.7 (31.8–217.6)	125.2 (24.1–222.9)	*t*-test: 0.55
Patient number	148	352	

*represents patients who can potentially benefit from the Irinotecan with $\Delta G < 2$ and † represents patients using irinotecan would lead to $\Delta G \geq 2$ based on the predicted probabilities of ΔG2 from bootstrapped models. *The markers correspond to the makers in* Fig. 3.

reduction up to 85 % of their original V_{10Gy}, which is the average bowel V_{10Gy} achievable when transitioning from Intensity-modulated radiation therapy (IMRT) to proton therapy, as reported by Berman et al. [14]. 23 (4.6 %) of these patients need a reduction factor below 50 %, which is the median small bowel V_{10Gy} reduction attainable when switching from IMRT or Three-dimensional conformal radiation therapy (3DCRT) to proton therapy, as indicated by Colaco et al. [15]. The feasible reduction ratio from proton to photon therapy is highly dependent on the relative positions of the tumour and small bowel, necessitating individual validation of feasibility.

Discussion

This study presents the influence of predictive features on treatment-related acute diarrhoea and models' accuracy for $\Delta G \geq 2$ based on the largest patient cohort as part of a prospective multicentre randomised trial. This larger cohort enhances the reliability of our predictive findings compared to previous studies. We also proposed a strategy to utilise the predictive models in guiding individualised radiotherapy plans, specifically focusing on minimising small bowel toxicity.

Challenges in model development using clinical data

Although an AUROC of 0.71 (95 % CI: 0.60–0.82) on the external dataset represents acceptable performance, it still offers only moderate accuracy for decision-support in routine practice. To further improve the model, we identified several challenges:

1 Human variability in contour delineation. The cases presented in Appendix A highlighted the challenges associated with manual contouring and underscored the utility of using automated segmentation tools for incorporating large datasets from various institutions into predictive models. We used the TotalSegmentator [9] to segment the small bowel in CT images to minimise inter-operator variability and account for the lack of contours. The inclusion of a diverse patient population, facilitated by the application of the TotalSegmentator, enhanced the robustness and reliability of our study's conclusions.

2 Lack of comprehensive clinical data for predictive models. The clinical-trial report is still unpublished. Our current dataset lacked detailed records on the administration of anti-diarrheal treatments to individual patients. According to the trial protocol, loperamide was recommended as a preventive treatment for diarrhoea. However, crucial details regarding loperamide administration, such as dosage and frequency, were not collected. Furthermore, for patients in arm B, who received irinotecan, a chemotherapeutic agent known to cause diarrhoea, it remained unclear whether additional anti-diarrheal measures were employed. Therefore, the toxicity grades in our dataset likely represent symptoms after mitigation. Patients who would have met criteria for grade-2 diarrhoea but achieved prompt relief through loperamide, bulking agents, or diet modification may have been recorded as lower-grade events, introducing outcome-label noise. This is consistent with the high false-positive rate observed in Arm B (Section 3.2). If clinicians were aware of the elevated risk of $\Delta G \geq 2$ in these patients and proactively implemented effective anti-diarrhoea treatment to mitigate small bowel toxicity [16], the model may have picked up patterns related to this management, contributing to the higher false-positive rate. This information gap is significant as the approach to

managing diarrhoea can substantially influence the outcome of acute toxicity. Several other factors known to affect the severity of diarrhoea were not captured in our dataset, including specific dietary modifications, genetic predispositions, the presence of diabetes, smoking habits, and the composition of the intestinal microbiota [17–20]. The development of accurate predictive models heavily relies on the availability of comprehensive and well-documented patient data. This necessitates the implementation of a meticulous recording system that requires inputs from multiple sources, including not only clinicians but also patients themselves.

3 *Development of model interpretation tools*. In this study, we employed a well-defined logistic regression model. However, the individual coefficients do not capture the interactions between variables. SHAP values provide a comprehensive measure of each feature's importance, accounting for both its individual effect and its synergistic or antagonistic relationships with other variables, ensuring a fair attribution of the model's output to each feature. Having a transparent model interpretation tool, such as SHAP, not only provides deeper insights into the model's decision-making process but also builds trust among clinicians and researchers, enabling them to make informed decisions based on the model's predictions.

Clinical application

In this study, the model was evaluated using an independent test set derived from the same multicentre clinical trial. Although both the training and test data originated from the same trial, the test cases were from different medical centres than those used for training. This provides a degree of external validity for the treatment regimen being studied. In future work, we aim to further validate the model using cohorts from additional institutions to strengthen its generalisability. We also plan to investigate alternative treatment strategies, such as intravenous 5-fluorouracil (5-FU), a commonly used alternative to oral capecitabine in rectal cancer chemoradiotherapy, to examine whether the observed dose–toxicity patterns and SHAP-based explanations remain consistent across different treatment protocols and clinical settings.

Previous studies have shown consistently that a low radiation dose to a large volume of the small bowel is predictive of acute diarrhoea in patients undergoing rectal cancer treatment and the correlation between dose-volume metrics at different dose levels, for example, V_{10Gy} and V_{15Gy}, is high [1,5,7]. In this study, we selected V_{10Gy} build the model, whilst V_{15Gy} being the commonly clinically accepted parameter [21] for interventions to minimise toxicity because proton therapy offers a significant difference at the 10 Gy dose levels, compared with 3D conformal radiotherapy/IMRT due to the dosimetric properties of proton beams [22].

Intensity-modulated proton therapy has shown its advantages in minimizing the dose to normal tissues [12,23,24], exploiting the steep falloff of the Bragg peak. However, it is important to note that proton therapy comes with higher initial costs compared to photon therapy [25]. Our model permits the identification of patients who could potentially see a greater benefit from proton therapy, based on their likelihood of experiencing reduced toxicity from a lower V_{10Gy}. This method of selection approach could serve as a validation of our predictive model and potentially offer a more personalized and effective treatment option for patients at high risk of treatment-induced toxicity, while also considering the cost-benefit aspect of employing proton therapy over traditional photon therapy.

The missing information related to acute diarrhoea, along with variability in the training data, contributes to the variability observed in our model's predictions. Enhancing the model with additional relevant parameters will be explored in the future for further development and clinic use.

Conclusion

We demonstrated prediction models for the development of acute diarrhoea toxicity due to chemoradiotherapy in the treatment of rectal cancer. We proposed using the models' predictions to inform clinical decisions, specifically in the context of individualising radiotherapy planning. By identifying patients at higher risk of experiencing significantly increased toxicity, this per-patient analysis we proposed can guide adjustments to treatment plans to minimise this risk. Our approach demonstrated the potential to understand treatment outcomes and personalise patient care.

Funding statement

This project is funded by UKRI Future Leaders Fellowship, No. MR/T040785/1 and the Radiation Research Unit at the Cancer Research UK City of London Centre Award C7893/A28990 and CRUK RRNPSF-Jan21 \100001 grant. Maria Hawkins is supported by the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre. Nicholas West is funded by a CRUK Radiation Research Network Infrastructure Award (RRNIA-Feb22\100003) and is supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC) (NIHR203331). Ane Appelt, Alexandra Gilbert, and David Sebag-Montefiore would like to acknowledge Cancer Research UK funding for the Leeds Radiotherapy Research Centre of Excellence (RadNet; C19942/A28832). The ARISTOTLE trial is funded by Cancer Research UK C19942/A10016. The views expressed are those of the author(s) and not necessarily those of the NHS, is also from United Kingdom, the NIHR or the Department of Health and Social Care.

Data sharing statement

The data that support the findings of this study are not publicly available due to ethical restrictions.

Patient consent statement

IRB approvals were obtained.

CRediT authorship contribution statement

Ying Zhang: Methodology, Formal analysis, Writing – original draft, Investigation. **Douglas Brand:** Conceptualization, Writing – review & editing. **Zhuoyan Shen:** Data curation, Writing – review & editing. **Mikael Simard:** Data curation, Writing – review & editing. **Sumeet Hindocha:** Data curation, Writing – review & editing. **Olivia Chohan:** Data curation. **Andre Lopes:** Writing – review & editing, Data curation. **Rubina Begum:** Conceptualization. **Nicholas West:** Data curation, Writing – review & editing. **Ane Appelt:** Conceptualization, Writing – review & editing. **Alexandra Gilbert:** Data curation, Writing – review & editing. **Elizabeth Miles:** Conceptualization, Writing – review & editing. **Tim Maughan:** Conceptualization, Writing – review & editing. **David Sebag-Montefiore:** Data curation, Writing – review & editing. **Maria A. Hawkins:** Writing – review & editing, Conceptualization, Methodology. **Charles-Antoine Collins Fekete:** Methodology, Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project is funded by UKRI Future Leaders Fellowship, No. MR/T040785/1 and the Radiation Research Unit at the Cancer Research UK City of London Centre Award C7893/A28990 and CRUK RRNPSF-Jan21\100001 grant. Maria Hawkins is supported by the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre. Nicholas West is funded by a CRUK Radiation Research Network Infrastructure Award (RRNIA-Feb22\100003) and is supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC) (NIHR203331). Ane Appelt, Alexandra Gilbert, and David Sebag-Montefiore would like to acknowledge Cancer Research UK funding for the Leeds Radiotherapy Research Centre of Excellence (RadNet; C19942/A28832). The ARISTOTLE trial is funded by Cancer Research UK C19942/A10016. The views expressed are those of the author(s) and not necessarily those of the NHS, is also from United Kingdom, the NIHR or the Department of Health and Social Care.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2025.111032.

References

[1] Reis T, Khazzaka E, Welzel G, Wenz F, Hofheinz RD, Mai S. Acute small-bowel toxicity during neoadjuvant combined radiochemotherapy in locally advanced rectal cancer: Determination of optimal dose-volume cut-off value predicting grade 2-3 diarrhoea. Radiat. Oncol. 2015;10. https://doi.org/10.1186/s13014-015-0336-5.

[2] Brændengen, Morten, Kjell M. Tveit, Åke Berglund, Elke Birkemeyer, Gunilla Frykholm, Lars Påhlman, Johan N. Wiig, Per Byström, Krzysztof Bujko, and Bengt Glimelius. 2008. "Randomized phase III study comparing preoperative radiotherapy with chemoradiotherapy in nonresectable rectal cancer." *Journal of Clinical Oncology* 26 (22). https://doi.org/10.1200/JCO.2007.15.3858.

[3] Robertson JM, Lockman D, Yan Di, Wallace M. The Dose-volume Relationship of Small Bowel Irradiation and Acute Grade 3 Diarrhea During Chemoradiotherapy for Rectal Cancer. Int. J. Radiat. Oncol. Biol. Phys. 2008;70. https://doi.org/10.1016/j.ijrobp.2007.06.066.

[4] Huang, Eng Yen, Yu Ming Wang, Shih Chen Chang, Shu Yu Liu, and Ming Chung Chou. 2021. "Rectal dose is the other dosimetric factor in addition to small bowel for prediction of acute diarrhea during postoperative whole-pelvic intensity-modulated radiotherapy in gynecologic patients." *Cancers* 13 (3). https://doi.org/10.3390/cancers13030497.

[5] Arbea L, Ramos L, Beunza J, Moreno M, Cambeiro M, Martinez-Monge R, et al. Dosimetric Predictors of Gastrointestinal Toxicity During Intensity Modulated Radiation Therapy Concomitant with Capecitabine and Oxaliplatin for locally Advanced Rectal Cancer. Int. J. Radiat. Oncol. Biol. Phys. 2012;84. https://doi.org/10.1016/j.ijrobp.2012.07.2126.

[6] Robertson JM, Söhn M, Yan Di. Predicting Grade 3 Acute Diarrhea During Radiation Therapy for Rectal Cancer using a Cutoff-Dose Logistic Regression Normal Tissue Complication Probability Model. Int. J. Radiat. Oncol. Biol. Phys. 2010;77. https://doi.org/10.1016/j.ijrobp.2009.04.048.

[7] Holyoake DLP, Partridge M, Hawkins MA. Systematic review and meta-analysis of small bowel dose–volume and acute toxicity in conventionally-fractionated rectal cancer radiotherapy. Radiother. Oncol. 2019;138. https://doi.org/10.1016/j.radonc.2019.05.001.

[8] Sebag-Montefiore, D., Adams, R., Gollins, S., Samuel, L. M., Glynne-Jones, R., Harte, R., ... & ARISTOTLE Trial Management Group. (2020). ARISTOTLE: a phase III trial comparing concurrent capecitabine with capecitabine and irinotecan (Ir) chemoradiation as preoperative treatment for MRI-defined locally advanced rectal cancer (LARC).

[9] Wasserthal, Jakob, Hanns Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, et al. 2023. "TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images." *Radiology: Artificial Intelligence* 5 (5). https://doi.org/10.1148/ryai.230024.

[10] Efron B, Tibshirani RJ. An introduction to the bootstrap. CRC Press; 1994.

[11] Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. Sci. Rep. 2022;12:5979. https://doi.org/10.1038/s41598-022-09954-8.

[12] Lundberg, Scott M., and Su In Lee. 2017. "A unified approach to interpreting model predictions." In *Advances in Neural Information Processing Systems*. Vol. 2017–December.

[13] Baumgartner W, Weiß P, Schindler H. A nonparametric test for the general two-sample problem. Biometrics 1998:1129–35.

[14] Berman AT, Both S, Sharkoski T, Goldrath K, Tochner Z, Apisarnthanarax S, et al. Proton reirradiation of recurrent rectal cancer: dosimetric comparison, toxicities, and preliminary outcomes. Int. J. Particle Therapy 2014;1:2–13. https://doi.org/10.1016/j.adro.2020.10.008.

[15] Colaco, Rovel J., Romaine Charles Nichols, Soon Huh, Nataliya Getman, Meng Wei Ho, Zuofeng Li, Christopher G. Morris, William M. Mendenhall, Nancy P. Mendenhall, and Bradford S. Hoppe. "Protons offer reduced bone marrow, small bowel, and urinary bladder exposure for patients receiving neoadjuvant radiotherapy for resectable rectal cancer." Journal of gastrointestinal oncology 5, 1 (2014): 3. https://doi.org/10.3978%2Fj.issn.2078-6891.2013.041.

[16] Pergolizzi, Stefano, Ernesto Maranzano, Verena De Angelis, Marco Lupattelli, Paolo Frata, Stefano Spagnesi, Maria Luisa Frisio, et al. 2013. "Diarrhoea in irradiated patients: A prospective multicentre observational study." *Digestive and Liver Disease* 45 (11). https://doi.org/10.1016/j.dld.2013.04.012.

[17] Nina NB, Pötter R, Spampinato S, Fokdal LU, Chargari C, Lindegaard JC, et al. Dose-volume effects and risk factors for late diarrhea in cervix cancer patients after radiochemotherapy with image guided adaptive brachytherapy in the EMBRACE i study. Int. J. Radiat. Oncol. Biol. Phys. 2021;109. https://doi.org/10.1016/j.ijrobp.2020.10.006.

[18] Eaton, Selina E., Justyna Kaczmarek, Daanish Mahmood, Anna M. McDiarmid, Alya N. Norarfan, Erin G. Scott, Chee Kin Then, Hailey Y. Tsui, and Anne E. Kiltie. 2022. "Exploiting dietary fibre and the gut microbiota in pelvic radiotherapy patients." https://doi.org/10.1038/s41416-022-01980-7.

[19] Liu, Meng Meng, Shu Ting Li, Yan Shu, and He Qin Zhan. 2017. "Probiotics for prevention of radiation-induced diarrhea: A meta-Analysis of randomized controlled trials." *PLoS ONE* 12 (6). https://doi.org/10.1371/journal.pone.0178870.

[20] Yi, Yuxi, Weiqing Lu, Lijun Shen, Yang Wu, and Zhen Zhang. 2023. "The gut microbiota as a booster for radiotherapy: novel insights into radio-protection and radiation injury." https://doi.org/10.1186/s40164-023-00410-5.

[21] Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. Int. J. Radiat. Oncol. Biol. Phys. 2010;76:S3–9. https://doi.org/10.1016/j.ijrobp.2009.09.040.

[22] Fok M, Toh S, Easow J, Fowler H, Clifford R, Parsons J, et al. Proton beam therapy in rectal cancer: a systematic review and meta-analysis. Surg. Oncol. 2021;38. https://doi.org/10.1016/j.suronc.2021.101638.

[23] Wolff, Hendrik Andreas, Daniela Melanie Wagner, Lena Christin Conradi, Steffen Hennies, Michael Ghadimi, Clemens Friedrich Hess, and Hans Christiansen. 2012. "Irradiation with protons for the individualized treatment of patients with locally advanced rectal cancer: A planning study with clinical implications." *Radiotherapy and Oncology* 102 (1). https://doi.org/10.1016/j.radonc.2011.10.018.

[24] Vaios, Eugene J., and Jennifer Y. Wo. 2020. "Proton beam radiotherapy for anal and rectal cancers." https://doi.org/10.21037/jgo.2019.04.03.

[25] Verma, Vivek, Chirag Shah, Jean Claude M. Rwigema, Timothy Solberg, Xiaofeng Zhu, and Charles B. Simone. 2016. "Cost-comparativeness of proton versus photon therapy." https://doi.org/10.21037/cco.2016.06.03.