

**Supplementary Table 4:** Concordance between lesions on baseline MRI (blue scale) and between variation of MRI lesions between pre- and post-treatment MRIs (red scale).

	Baseline MRI									
	Ulcer	Edema	Qualitative variables				Quantitative variables			
			Qualitative contrast enhancement	Qualitative diffusion-weighted hyperintensity	Bowel thickness > 3 mm	Fat wrapping	Enlarged lymph nodes	ADC	RCE	Bowel thickness (mm)
Ulcer		0.20 ± 0.035 (69.5%)	0.13 ± 0.028 (67.5%)	0.16 ± 0.036 (67.8%)	0.10 ± 0.024 (66.5%)	0.24 ± 0.049 (61.3%)	0.19 ± 0.038 (54.1%)	ES = -0.42 1.40 vs 1.54 p = 0.0008	ES = 0.25 168 vs 151 p = 0.036	ES = 0.56 7.3 vs 6.1 p < 0.0001
Edema	0.54 ± 0.061 (86.9%)		0.53 ± 0.050 (94.7%)	0.42 ± 0.053 (91.5%)	0.53 ± 0.047 (95.0%)	0.07 ± 0.025 (50.0%)	0.03 ± 0.030 (30.8%)	ES = -0.90 1.43 vs 1.73 p = 0.0025	ES = 1.27 170 vs 85 p = 0.0002	ES = 1.15 7.0 vs 4.6 p < 0.0001
Qualitative contrast enhancement	0.53 ± 0.060 (86.3%)	0.93 ± 0.056 (97.8%)		0.60 ± 0.050 (95.2%)	0.77 ± 0.052 (98.3%)	0.05 ± 0.019 (48.2%)	0.01 ± 0.013 (27.8%)	ES = -1.49 1.43 vs 1.90 p < 0.0001	ES = 2.00 168 vs 39 p < 0.0001	ES = 1.52 7.0 vs 3.8 p < 0.0001
Qualitative diffusion-weighted hyperintensity	0.57 ± 0.063 (87.8%)	0.89 ± 0.057 (98.2%)	0.88 ± 0.056 (96.5%)		0.51 ± 0.047 (94.7%)	0.06 ± 0.026 (49.7%)	0.01 ± 0.014 (28.4%)	ES = -1.36 1.42 vs 1.84 p < 0.0001	ES = 1.06 168 vs 97 p = 0.0004	ES = 0.79 7.0 vs 5.3 p = 0.0004
Bowel thickness > 3 mm	0.50 ± 0.058 (85.8%)	0.94 ± 0.055 (96.7%)	0.95 ± 0.054 (98.8%)	0.88 ± 0.055 (96.5%)		0.04 ± 0.017 (47.9%)	0.01 ± 0.011 (27.1%)	ES = -1.58 1.43 vs 1.96 p < 0.0001	ES = 2.24 167 vs 24 p < 0.0001	ES = 1.95 7.0 vs 3.0 p < 0.0001
Fat wrapping	0.36 ± 0.085 (87.1%)	0.82 ± 0.0805 (94.8%)	0.76 ± 0.080 (93.5%)	0.80 ± 0.081 (94.1%)	0.79 ± 0.078 (94.4%)		0.01 ± 0.014 (29.4%)	ES = -0.38 1.39 vs 1.52 p = 0.0003	ES = 0.09 166 vs 159 p = 0.36	ES = 0.35 7.3 vs 6.5 p = 0.0008
Enlarged lymph nodes	0.50 ± 0.115 (86.1%)	0.58 ± 0.10 (89.0%)	0.52 ± 0.10 (87.7%)	0.44 ± 0.10 (87.1%)	0.58 ± 0.10 (89.2%)	0.56 ± 0.12 (92.4%)		ES = -0.26 1.39 vs 1.48 p = 0.011	ES = 0.03 164 vs 161 p = 0.78	ES = 0.10 7.0 vs 6.8 p = 0.29
ADC	ES = 0.84 0.25 vs 0.00 p < 0.0001	ES = 1.11 0.37 vs 0.01 p < 0.0001	ES = 1.04 0.37 vs 0.02 p < 0.0001	ES = 1.28 0.41 vs 0.02 p < 0.0001	ES = 1.04 0.38 vs 0.02 p < 0.0001	ES = 1.18 0.41 vs 0.02 p < 0.0001	ES = 0.90 0.23 vs 0.00 p = 0.013		q = -0.04	q = -0.24
RCE	ES = 0.84 70 vs 2 p < 0.0001	ES = 1.93 145 vs 6 p < 0.0001	ES = 1.95 151 vs 5 p < 0.0001	ES = 1.49 128 vs 7 p < 0.0001	ES = 1.90 149 vs 5 p < 0.0001	ES = 1.19 106 vs 10 p = 0.002	ES = 0.66 51 vs 0 p = 0.07	q = -0.28		q = 0.13
Bowel thickness (mm)	ES = 1.08 1.9 vs 0.1 p < 0.0001	ES = 1.66 3.2 vs 0.3 p < 0.0001	ES = 1.52 3.1 vs 0.3 p < 0.0001	ES = 1.34 2.8 vs 0.3 p < 0.0001	ES = 1.66 3.2 vs 0.3 p < 0.0001	ES = 1.22 2.5 vs 0.2 p < 0.0001	ES = 1.15 2.1 vs 0.2 p = 0.0013	q = -0.35	q = 0.38	

For the relationship between two qualitative parameters, the results are expressed as  $\kappa$  ± standard error and agreement (in round brackets). For the relationship between one qualitative and one quantitative parameters, the results are expressed as effect size (ES), mean values with or without the presence of the lesion and p-values. For the relationship between two quantitative parameters, the results are expressed as coefficient of correlation (q).

Concordance ( $\kappa$ )	0.00 – 0.19	0.20 – 0.39	0.4 – 0.59	0.60 – 0.79	≥ 0.80	Concordance ( $\kappa$ )	0.00 – 0.19	0.20 – 0.39	0.4 – 0.59	0.60 – 0.79	≥ 0.80
Effect size (ES)	0.00 – 0.19	0.20 – 0.49	0.50 – 0.79	0.80 – 1.59	≥ 1.60	Effect size (ES)	0.00 – 0.19	0.20 – 0.49	0.50 – 0.79	0.80 – 1.59	≥ 1.60