Radiology

Primary Rectal Cancer:

Repeatability of Global and Local-Regional MR Imaging Texture Features[1]

Sofia Gourtsoyianni, PhD
Georgia Doumou, MRes
Davide Prezzi, FRCR
Benjamin Taylor, FRCR
J. James Stirling, MSc
N. Jane Taylor, PhD
Musib Siddique, PhD
Gary J. R. Cook, MD, FRCR, FRCP
Robert Glynne-Jones, MD, FRCR, FRCP
Vicky Goh, MD, FRCR

Purpose: To assess the day-to-day repeatability of global and local-regional magnetic resonance (MR) imaging texture features derived from primary rectal cancer.

Materials and Methods: After ethical approval and patient informed consent were obtained, two pretreatment T2-weighted axial MR imaging studies performed prospectively with the same imaging unit on 2 consecutive days in 14 patients with rectal cancer (11 men [mean age, 61.7 years], three women [mean age, 70.0 years]) were analyzed to extract *(a)* global first-order statistical histogram and model-based fractal features reflecting the whole-tumor voxel intensity histogram distribution and repeating patterns, respectively, without spatial information and *(b)* local-regional second-order and high-order statistical texture features reflecting the intensity and spatial interrelationships between adjacent in-plane or multiplanar voxels or regions, respectively. Repeatability was assessed for 46 texture features, and mean difference, 95% limits of agreement, within-subject coefficient of variation (wCV), and repeatability coefficient *(r)* were recorded.

Results: Repeatability was better for global parameters than for most local-regional parameters. In particular, histogram mean, median, and entropy, fractal dimension mean and standard deviation, and second-order entropy, homogeneity, difference entropy, and inverse difference moment demonstrated good repeatability, with narrow limits of agreement and wCVs of 10% or lower. Repeatability was poorest for the following high-order gray-level run-length (GLRL) gray-level zone size matrix (GLZSM) and neighborhood gray-tone difference matrix (NGTDM) parameters: GLRL intensity variability, GLZSM short-zone emphasis, GLZSM intensity nonuniformity, GLZSM intensity variability, GLZSM size zone variability, and NGTDM complexity, demonstrating wider agreement limits and wCVs of 50% or greater.

Conclusion: MR imaging repeatability is better for global texture parameters than for local-regional texture parameters, indicating that global texture parameters should be sufficiently robust for clinical practice.

Published under a CC BY 4.0 license.

Online supplemental material is available for this article.

It is now recognized that tumors may demonstrate considerable biologic heterogeneity that influences their clinical outcome (1). Heterogeneity is also apparent phenotypically and can be appreciated in medical images (2) where traditionally radiology descriptors such as heterogeneous signal intensity or enhancement, mixed signal pattern, presence of areas of necrosis, hemorrhage, or calcification have been used to capture such morphologic information. Descriptive approaches have a clinical role but have limitations, and there has been increasing interest in alternative,

quantitative method to portray this information, facilitated by the availability of texture analysis software platforms in clinical and research practice (3–6).

Results of a number of magnetic resonance (MR) imaging–based studies have now suggested that texture analysis may augment conventional imaging in cancer therapy response assessment (7–10) and may have a role as a prognostic and predictive biomarker, either alone or combined with clinical and genomic information (7,11). To date, there have been few studies assessing the robustness of such analysis (12–16). Image heterogeneity may be quantified by a number of postprocessing methods that address the variation of gray levels within a volume of interest, either on a global (whole-tumor) or local-regional (intratumoral, in-plane, or multiplanar) scale (Table 1, Tables E1–E7 [online]). These include (a) global model-based parameters such as fractal dimension and lacunarity, (b) global first-order statistical parameters derived from the gray-level intensity histograms (eg, mean, median, skewness, kurtosis, first-order entropy, and energy, also known as uniformity), (c) local second-order statistical parameters derived by using co-occurrence matrixes that reflect the spatial and signal intensity interrelationships between adjacent in-plane voxels (eg, contrast, homogeneity, second-order entropy, and energy), and (d) local-regional high-order parameters that examine the relationship between neighboring voxels in adjacent imaging planes (these include NGTDM parameters [eg, coarseness, contrast, busyness, complexity], which describe the dynamic range of intensities at a local level; run-length parameters, which describe the variation in signal intensity

level across longitudinal voxels; and variations in zones or regions of intensity). The repeatability of these different texture parameters remains unknown in rectal cancer, there being no published studies with MR imaging to date (to our knowledge). Repeatability, an indication of measurement error, is essential information for clinical trials and future clinical practice; for example, if these quantitative parameters are used to augment therapy assessment. Thus, the aim of our study was to assess the day-to-day repeatability of global and local-regional MR imaging texture features derived from primary rectal cancer.

Materials and Methods

Patients

Institutional review board approval and informed consent were obtained for this study. Adults with locally advanced rectal cancer (stage $\geq$ T3N-positive or T4N0) who were scheduled to undergo chemotherapy and radiation therapy (45 Gy in 25 fractions with concomitant capecitabine 850 mg/m^2) prior to surgery were recruited prospectively for a study investigating response biomarkers to preoperative chemotherapy

Advances in Knowledge

■ Several texture parameters demonstrated good repeatability, with narrow limits of agreement and within-subject coefficients of variation (wCVs) of 10% or lower.

■ Global parameters that provided a global measure of signal distribution and demonstrated good repeatability included (a) histogram parameters (first-order histogram mean, median, and entropy) and (b) model-based fractal dimension mean and standard deviation.

■ Local parameters within a defined section within the lesion that reflected the intratumoral intensity and spatial interrelationships between adjacent in-plane voxels and demonstrated good repeatability included second-order gray-level co-occurrence matrix entropy, homogeneity, difference entropy, and inverse difference moment.

■ Repeatability was poorest for high-order texture parameters that provided information about intratumoral local-regional intensity and spatial interrelationships between multiplanar voxels or regions (in particular, several neighborhood gray-tone difference matrix and gray-level zone size parameters showed wCVs of 50% or greater).

Implication for Patient Care

■ The day-to-day repeatability of MR imaging global whole-tumor texture parameters is excellent for rectal cancer and opens the possibility of future clinical usage of whole-tumor MR imaging texture analysis in view of its potential prognostic information.

Table 1

Summary of Global and Local-Regional Features Extracted for Each Tumor

Type of Measure and Measure	Scale	Method	Parameters
First order			
Histogram statistics	Global	Describes the distribution of voxel signal intensity values without spatial information.	Mean, median, skewness, kurtosis, entropy, and energy
Second order			
GLCM	Local	Describes the statistical interrelationships between voxels with similar or dissimilar signal intensity values within an imaging plane.	Entropy, homogeneity, energy, contrast, autocorrelation, cluster shade, cluster prominence, difference entropy, difference variance, dissimilarity, inverse difference moment, maximum probability, sum average, sum entropy, and sum variance
Gray-level difference matrix	Regional	Describes the difference in signal intensity and displacement between two voxels within an imaging plane.	Mean, entropy, variance, and contrast
High order			
NGTDM	Local	Describes the signal intensity and spatial interrelationship between neighboring voxels between adjacent image planes.	Coarseness, contrast, busyness, complexity, and texture strength
Gray-level run length	Regional	Describes the signal intensity and spatial interrelationship along a longitudinal run of neighboring voxels.	Short-run emphasis, long-run emphasis, gray-level nonuniformity, run-length nonuniformity, intensity variability, and run-length variability
GLZSM	Regional	Describes the distribution of similar and dissimilar regions with intensity variations	Short zone emphasis, long zone emphasis, intensity nonuniformity, zone length nonuniformity, intensity variability, and size zone variability
Model based			
Fractal analysis	Global	Describes repetitive patterns within an image extracted by using filter grids.	Fractal dimension mean, fractal dimension standard deviation, lacunarity, and Hurst exponent

Note.—GLCM = gray-level co-occurrence matrix, GLZSM = gray-level zone size matrix (Figure), NGTDM = neighborhood gray-tone difference matrix.

and radiation therapy from May 2009 through January 2012. Exclusion criteria were contraindications to contrast material–enhanced MR imaging and metastatic disease at staging. Seventeen patients initially consented to undergo two pretreatment MR imaging examinations on consecutive days. Two patients subsequently declined a second baseline MR imaging examination, and one had hepatic metastatic disease, leaving a study group of 14 patients (11 men [mean age, 61.7 years; range, 52–79 years] and three women [mean age, 70.0 years; range, 60–78 years]).
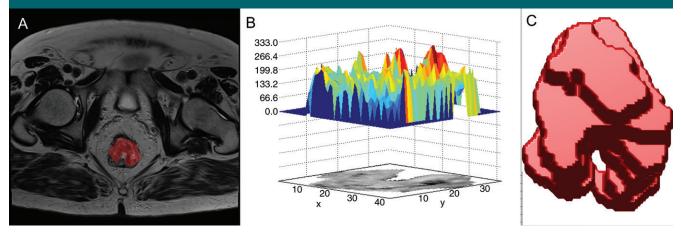
MR Imaging and Image Analysis

The pretreatment baseline MR imaging examinations were performed on consecutive days to allow assessment of day-to-day repeatability by using a 1.5-T system (Avanto and Symphony; Siemens, Erlangen, Germany). Both MR imaging examinations in each patient were performed on the same imaging unit by the same operator. No distention of the rectum was performed, and no rectal contrast material was administered. The T2-weighted axial turbo spin-echo pelvic MR images (repetition time msec/echo time msec, 4000/100; field of view, 263 × 350; matrix, 288 × 512 mm; section thickness, 6 mm; intersection gap, 1.8 mm) were selected and exported for further analysis. Texture analysis was performed by using an in-house program based on Matlab (Mathworks, Natick, Mass) to extract global whole-tumor and local-regional intratumoral features by using first-order, second-order, and high-order statistical methods and model-based methods (Table 1, Tables E1–E7 [online]). Image analysis was performed separately by two readers (G.D. [reader 1] and S.G. [reader 2], with 1 year and > 10 years of experience in MR imaging, respectively).

A region of interest was defined around the visualized tumor border by each reader on each axial image to segment a tumor volume for analysis (Figure), and subsequent analysis was automated within the software platform. A medium smoothing filter and 32-bin width were applied for quantization by the software. Global and local-regional texture features as listed in Table 1 and Tables E1–E7 (online) were extracted automatically for the defined tumor volume. This process was repeated for the second baseline MR imaging study, and the same global whole-tumor and local-regional intratumoral texture features were extracted for the defined tumor volume.

Statistical Analysis

To assess repeatability, the mean difference and 95% limits of agreement (with

Segmentation of a rectal tumor for heterogeneity analysis: An example of tumor delineation on, *A,* a single axial MR image, *B,* a corresponding surface histogram plot, and, *C,* a subsequent whole-tumor composite volume on which analysis was performed.

95% confidence intervals [CIs] for the upper and lower limits of agreement) were determined for each of the texture parameters (17) for each observer. The mean difference reflects the difference between two separate individual measurements, while the limits of agreement provide the 95% CIs for the difference between individual measurements between different studies.

The within-subject coefficient of variation (wCV), representing the within-subject variability of parameters and expressed as a percentage, was calculated by dividing the within-subject standard deviation by the group mean as follows:

$$\mathrm{wCV} = \frac{\sqrt{\frac{\Sigma d^2}{n}} / \sqrt{2}}{\mathrm{mean}},$$

where d is the difference and n is the number of subjects. wCVs of greater than 50% were considered to indicate unreliability for clinical implementation.

The repeatability coefficient representing the precision of repeated measures was calculated as follows:

$$\sqrt{2} \cdot 1.96 \cdot \sqrt{\frac{\Sigma d^2}{n}} / \sqrt{2},$$

where d is the difference and n is the number of subjects. This is a 95% CI:

The difference between two measurements for the same subject will be less than this value in 95% of cases.

Results

There were 10 T3N1 tumors, three T3N2 tumors, and one T4N0 tumor; mean tumor length was 50 mm (range, 30–85 mm). The mean differences, 95% limits of agreement, wCVs, and repeatability coefficients for the 46 image texture parameters are summarized in Tables 2–8 for both observers. Day-to-day repeatability was variable depending on the class and type of parameter.

Repeatability was generally better for whole-tumor global parameters (Tables 2, 3) than for intratumoral local-regional parameters, with the exception of first-order skewness, kurtosis, fractal lacunarity, and Hurst exponent. In particular, first-order mean, median, and entropy and fractal dimension mean and standard deviation demonstrated narrow limits of agreement, with low within-subject variability of 10% or less. The repeatability of first-order energy was also acceptable, with the within-subject variability ranging from 10% to 15%.

For second-order in-plane local-regional parameters (Tables 4, 5), only GLCM entropy, GLCM homogeneity, GLCM difference entropy, and GLCM inverse difference moment demonstrated narrow limits of agreement with low within-subject variability of 10% or less. Cluster shade, cluster prominence, and sum entropy demonstrated only moderate limits of agreement and within-subject variability of 30% or more for at least one observer.

Repeatability was poorest for high-order local-regional parameters (Tables 6–8), which reflects the interrelationship between multiplanar voxels. GLZSM and NGTDM parameters, with the exception of NGTDM coarseness, demonstrated poorer limits of agreement and within-subject variability of 30% or greater for at least one observer. In particular, NGTDM complexity, gray-level run-length intensity variability, GLZSM short-zone emphasis, GLZSM intensity nonuniformity, GLZSM intensity variability, and GLZSM size zone variability demonstrated within-subject variability of 50% or greater.

Discussion

Low measurement error is essential for any quantitative technique to be translated into clinical practice. In our study, we found that global methods that reflect the voxel distribution within the whole tumor demonstrated higher repeatability than did local-regional methods that describe the intervoxel relationships on a local or regional level. Specifically, global first-order

Table 2

Repeatability between Two Baseline MR Imaging Examinations for Global First-Order Statistical Histogram Parameters

Parameter and Reader	Mean Difference	95% Limits of Agreement	95% CI for Lower Limits	95% CI for Upper Limits	wCV (%)	Repeatability Coefficient
Mean						
Reader 1	+1.37	−51.38, +48.65	−72.94, −22.35	+25.09, +75.67	8.85	47.28
Reader 2	−4.07	−55.68, +47.53	−81.49, −29.88	+21.73, +73.34	9.05	49.36
Median						
Reader 1	−1.46	−49.28, +46.36	−73.19, −25.37	+22.44, +70.26	8.69	45.22
Reader 2	−4.21	−54.54, +46.11	−79.71, −29.38	+20.95, +71.28	9.12	48.21
Skewness						
Reader 1	−0.092	−0.74, +0.55	−1.06, −0.41	+0.23, +0.88	25.23	0.64
Reader 2	−0.11	−1.00, +0.77	−1.44, −0.56	+0.33, +1.22	27.67	0.87
Kurtosis						
Reader 1	−0.49	−3.27, +2.28	−4.66, −1.88	+0.89, +3.67	17.73	2.79
Reader 2	−0.43	−5.92, +5.07	−8.67, −3.17	+2.32, +7.81	26.50	5.25
Entropy						
Reader 1	−0.025	−0.46, +0.41	−0.67, −0.24	+0.19, +0.62	2.97	0.41
Reader 2	−0.22	−0.38, +0.34	−0.56, −0.20	+0.16, +0.52	2.40	0.34
Uniformity						
Reader 1	+0.0002	−0.0035, +0.0039	−0.0053, −0.0016	+0.0021, +0.0057	14.82	0.0035
Reader 2	+0.0003	−0.002, +0.003	−0.004, +0.001	+0.002, +0.004	12.29	0.0025

Table 3

Repeatability between Two Baseline MR Imaging Examinations for Global Fractal Parameters

Parameter and Reader	Mean Difference	95% Limits of Agreement	95% CI for Lower Limits	95% CI for Upper Limits	wCV (%)	Repeatability Coefficient
Fractal dimension mean						
Reader 1	+0.013	−0.10, +0.13	−0.16, −0.045	+0.072, +0.19	2.17	0.11
Reader 2	+0.014	−0.044, +0.072	−0.073, −0.015	+0.043, +0.10	1.19	0.061
Fractal dimension standard deviation						
Reader 1	−0.00045	−0.069, +0.069	−0.10, −0.035	+0.034, +0.10	10.80	0.065
Reader 2	−0.0036	−0.051, +0.043	−0.074, −0.027	+0.019, +0.067	6.60	0.045
Lacunarity						
Reader 1	−0.00054	−0.012, +0.011	−0.017, −0.0062	+0.0051, +0.016	27.90	0.011
Reader 2	+0.0001	−0.007, +0.007	−0.011, −0.004	+0.004, +0.011	21.03	0.0069
Hurst exponent						
Reader 1	−0.013	−0.13, +0.10	−0.19, −0.07	+0.045, +0.16	35.5	0.11
Reader 2	−0.0103	−0.077, +0.056	−0.11, −0.043	+0.023, +0.089	5.81	0.066

mean, median, and entropy and fractal dimension mean and standard deviation demonstrated narrow limits of agreement, with low within-subject variability of 10% or less. First-order energy was also acceptable, with the within-subject variability ranging from 10% to 15% and thus being sufficiently robust for clinical practice. Of the local-regional parameters, gray-level difference matrix entropy, GLCM entropy, GLCM homogeneity, GLCM difference entropy, and GLCM inverse difference moment demonstrated the narrowest limits of agreement, from −0.29 to +0.29, −0.70 to +0.66, −0.099 to +0.10, −0.29 to +0.29, and −0.099 to +0.101, respectively.

Repeatability was poorest for high-order local-regional parameters. GLZSM and NGTDM parameters, with the exception of NGTDM coarseness, were poorly reproducible across the two MR imaging examinations. Although our observers had different levels of MR imaging experience, the level of repeatability was consistent for most parameters. Discrepancies, where present, were primarily related to high-order local-regional parameters—for example, NGTDM complexity, busyness, and contrast; gray-level run-length intensity and gray-level variability; and GLZSM

Table 4

Repeatability between Two Baseline MR Imaging Examinations for Local-Regional Second-Order Gray-Level Co-occurrence Matrix Texture Parameters

Parameter and Reader	Mean Difference	95% Limits of Agreement	95% CI for Lower Limits	95% CI for Upper Limits	wCV (%)	Repeatability Coefficient
Entropy						
Reader 1	−0.022	−0.70, +0.66	−1.04, −0.36	+0.32, +0.99	5.73	0.64
Reader 2	+0.036	−0.58, +0.65	−0.88, −0.27	+0.34, +0.95	5.40	0.58
Homogeneity						
Reader 1	+0.00086	−0.099, +0.10	−0.15, −0.049	+-0.051, +0.15	5.87	0.094
Reader 2	−0.012	−0.068, +0.044	−0.096, −0.040	+0.016, +0.072	3.38	0.058
Energy						
Reader 1	−0.0002	−0.031, +0.030	−0.046, −0.015	+0.015, 0.046	27.63	0.028
Reader 2	−0.0042	−0.035, +0.027	−0.050, −0.019	+0.011, +0.042	21.18	0.030
Contrast						
Reader 1	−0.075	−3.66, +3.51	−5.46, −1.87	+1.72, +5.31	21.00	3.39
Reader 2	+0.083	−3.34, +3.51	−5.06, −1.63	+1.79, +5.22	24.35	3.24
Autocorrelation						
Reader 1	−3.62	−74.27, +67.03	−109.60, −38.94	+31.69, +102.35	24.64	67.05
Reader 2	+3.65	−55.60, +62.91	−85.23, −25.97	+33.28, +92.54	23.30	56.38
Cluster shade						
Reader 1	−31.56	−214.70, +151.57	−306.27, −123.13	+60.00, +243.15	32.38	183.55
Reader 2	−24.23	−338.00, +289.54	−494.89, −181.12	+132.64, +446.43	43.05	299.89
Cluster prominence						
Reader 1	−1272.84	−8714.67, +6168.98	−12435.78, −4993.54	+2447.86, +9890.09	36.29	7452.37
Reader 2	+32.85	−1916.39, +1982.1	−2891.08, −941.72	+1007.42, +2956.78	52.78	1840.67
Difference entropy						
Reader 1	+0.0034	-0.29, +0.29	−0.44, −0.14	+0.15, +0.45	6.73	0.28
Reader 2	+0.046	−0.19, +0.28	−0.30, −0.069	+0.16, +0.39	6.37	0.24
Difference variance						
Reader 1	−0.075	−3.67, +3.51	−5.46, −1.87	+1.72, +5.31	21.00	3.39
Reader 2	−0.0022	−0.62, +0.62	−0.93, −0.31	+0.31, +0.93	13.87	0.58
Dissimilarity						
Reader 1	−0.0022	−0.62, +0.62	−0.93, −0.31	+0.31, +0.93	13.7	0.58
Reader 2	+0.071	−0.29, −0.44	−0.48, −0.11	+0.26, +0.61	11.14	0.37
Inverse difference moment						
Reader 1	+0.00086	−0.099, +0.101	−0.15, −0.049	+0.051, +0.15	5.87	0.095
Reader 2	−0.012	−0.068, +0.044	−0.096, −0.040	+0.016, +0.071	3.38	0.058
Maximum probability						
Reader 1	+0.00014	−0.058, +0.058	−0.087, −0.029	+0.029, +0.087	22.1	0.055
Reader 2	−0.0085	−0.068, +0.051	−0.097, −0.038	+0.021, +0.080	18.60	0.058
Sum average						
Reader 1	−0.43	−7.44, +6.58	−10.96, −3.94	+3.08, +10.09	13.52	6.67
Reader 2	+1.58	−5.51, +8.67	−9.06, +1.97	+5.12, +12.21	14.45	7.37
Sum entropy						
Reader 1	−0.018	−0.41, +0.37	−0.59, −0.21	+0.18, +0.56	4.57	0.37
Reader 2	+0.017	−0.36, +0.40	−0.56, −0.17	+0.20, +0.59	23.54	1.97
Sum variance						
Reader 1	−10.77	−245.83, +224.29	−363.36, −128.29	+106.75, +341.83	26.79	222.83
Reader 2	48.87	−179.59, +277.34	−293.83, −65.35	+163.09, +391.57	31.54	235.90

short-zone emphasis. This is most likely to be related to the differences in voxels included in the volume of interest delineation by the observers.

There have been few studies investigating the repeatability of both global and local-regional texture parameters, and none have been published for MR imaging, to our knowledge. Fried et al (12) reported an average correlation coefficient of 0.67 for contrast-enhanced computed tomography (CT) of stage III

Table 5

Repeatability between Two Baseline MR Imaging Examinations for Local-Regional Second-Order Gray-Level Difference Matrix Texture Parameters

Parameter and Reader	Mean Difference	95% Limits of Agreement	95% CI for Lower Limits	95% CI for Upper Limits	wCV (%)	Repeatability Coefficient
Mean						
Reader 1	−0.0027	−0.62, +0.62	−0.94, −0.31	+0.31, +0.93	13.89	0.59
Reader 2	+1.33	−0.40, +0.43	−0.61, −0.19	+0.22, +0.64	10.70	0.39
Entropy						
Reader 1	+0.0034	−0.29, +0.29	−0.44, −0.14	+0.15, +0.45	6.76	0.28
Reader 2	+0.007	−0.21, +0.22	−0.31, −0.098	+0.11, +0.33	5.16	0.20
Variance						
Reader 1	−0.073	−1.52, +1.37	−2.24, −0.79	+0.65, +2.09	17.88	1.37
Reader 2	+0.008	−1.79, +1.81	−2.69, −0.89	+0.91, +2.71	24.88	1.70
Contrast						
Reader 1	−0.10	−0.38, +0.36	−5.66, −1.95	+1.75, +5.45	21.50	3.50
Reader 2	+0.055	−3.32, +3.43	−5.02, −1.64	+1.75, +5.13	23.88	3.19

Table 6

Repeatability between Two Baseline MR Imaging Examinations for Local-Regional High-Order Neighborhood Gray-Tone Difference Matrix Texture Parameters

Parameter and Reader	Mean Difference	95% Limits of Agreement	95% CI for Lower Limits	95% CI for Upper Limits	wCV (%)	Repeatability Coefficient
Coarseness						
Reader 1	+0.034	−47.39, +47.46	−71.12, −23.68	+23.75, +71.18	19.43	44.76
Reader 2	+0.13	−15.59, +15.86	−23.46, −7.73	+7.99, +23.72	9.08	14.34
Contrast						
Reader 1	+0.00043	−0.0011, +0.0099	−0.016, −0.0056	+0.0047, +0.015	34.70	0.0099
Reader 2	−0.0053	−0.0089, +0.0079	−0.013, −0.0047	+0.0037, +0.012	45.25	0.008
Busyness						
Reader 1	−0.090	−2.25, +2.07	−3.33, −1.17	+0.99, +3.15	48.42	2.05
Reader 2	+0.15	−1.07, +1.36	−1.68, −0.46	+0.75, +1.97	26.71	1.18
Complexity						
Reader 1	−0.00092	−0.022, +0.020	−0.033, −0.011	+0.0098, +0.031	85.10	0.02
Reader 2	−0.00014	−0.00047, +0.00019	−0.00066, −0.00031	+0.00002, +0.00031	19.81	0.004
Texture strength						
Reader 1	−0.16	−1.64, +1.32	−2.37, −0.89	+0.58, +2.05	32.3	1.43
Reader 2	−0.19	−0.69, +0.29	−0.94, −0.44	+0.049, +0.55	38.3	0.61

non–small cell lung cancer (NSCLC) in 13 patients, with only 15 (23%) of the 66 features studied (first-order [histogram], second-order [GLCM] and high-order [NGTDM] statistical parameters, as well as Laplacian of Gaussian filtration metrics) showing a correlation coefficient greater than 0.9 (12). Similarly, Balagurunathan et al (13) reported that 66 (30.9%) texture features had a correlation coefficient greater than 0.9 in 32 patients with NSCLC undergoing noncontrast CT. Sanghera et al (14) found

that fractal dimension and lacunarity features in perfusion CT blood flow maps demonstrated coefficients of variation of less than 7.35% between studies. A further study of fluorine 18 fluorodeoxyglucose positron emission tomography statistic-based textural features in 16 patients with esophageal cancer showed variable reliability, with intraclass correlation coefficients ranging from 0.59 to 0.99 (15). Our findings for T2-weighted MR imaging demonstrate a similar pattern and degree of repeatability.

The use of more than one statistical method to evaluate repeatability reflects the fact that they provide complementary information regarding interstudy agreement. In our study, we selected Bland-Altman 95% limits of agreement, wCVs, and the repeatability coefficient, as they provide a measure of the variability of the parameters itself, a measure of the variability for individuals within the study group, and the degree of measured parameter change that is likely to be beyond day-to-day

Table 7

Repeatability between Two Baseline MR Imaging Examinations for Local-Regional High-Order Gray-Level Run-Length Texture Parameters

Parameter and Reader	Mean Difference	95% Limits of Agreement	95% CI for Lower Limits	95% CI for Upper Limits	wCV (%)	Repeatability Coefficient
Short-run emphasis						
Reader 1	−0.0064	−0.095, +0.082	−0.14, −0.051	+0.038, +0.13	25.49	0.084
Reader 2	−0.0015	−0.012, +0.0094	−0.018, −0.0069	+0.0039, +0.015	33.44	0.011
Long-run emphasis						
Reader 1	−5.17	−94.72, +105.06	−144.67, −44.77	+55.11, +155.01	22.47	94.81
Reader 2	+28.46	−69.84, +126.76	−118.99, −20.69	+77.61, +175.92	19.49	108.23
Gray-level nonuniformity						
Reader 1	+118.98	−2134.84, +1896.89	−3142.83, −1126.85	+888.89, +2904.87	18.64	1916.63
Reader 2	+520.83	−2109.78, +3151.44	−3425.16, −794.39	+1836.07, +4466.82	26.68	2683.98
Run-length nonuniformity						
Reader 1	−30.38	−471.92, +411.12	−692.69, −251.15	+190.35, +631.89	16.94	420.91
Reader 2	+37.80	−1782.69, +1858.29	−2692.99, −872.39	+947.99, +2768.59	22.49	1719.63
Intensity variability ($\times 10^{-3}$)						
Reader 1	−15.94	−211.83, +179.95	−309.77, −113.88	+81.99, +277.89	23.05	187.48
Reader 2	+64.87	−278.21, +407.95	−449.76, −106.66	+236.39, +579.49	51.03	347.81
Run-length variability ($\times 10^{-3}$)						
Reader 1	−3.93	−52.90, +45.05	−77.39, −28.41	+20.56, +69.54	21.58	46.85
Reader 2	+12.39	−31.98, +56.78	−54.17, −9.79	+34.59, +78.97	21.87	48.42

Table 8

Repeatability between Two Baseline MR Imaging Examinations for Local-Regional High-Order GLZSM Texture Parameters

Parameter and Reader	Mean Difference	95% Limits of Agreement	95% CI for Lower Limits	95% CI for Upper Limits	wCV (%)	Repeatability Coefficient
Short zone emphasis						
Reader 1	−0.014	−0.11, +0.082	−0.16, −0.062	+0.034, +0.13	90.25	0.094
Reader 2	−0.0026	−0.042, +0.037	−0.062, −0.023	+0.017, +0.057	43.44	0.038
Long zone emphasis						
Reader 1	+10.64	−108.60, +129.89	−168.23, −48.97	+70.26, +189.51	35.61	114.45
Reader 2	+13.83	−72.91, +100.57	−116.28, −29.57	+57.20, +143.94	23.13	86.23
Intensity nonuniformity						
Reader 1	−7.85	−72.36, +88.07	−112.47, −32.25	+47.96, +128.18	61.27	77.25
Reader 2	+13.82	−96.99, +124.62	−152.39, −41.58	+69.22, +180.03	71.32	108.01
Zone length nonuniformity						
Reader 1	−0.75	−17.49, +16.00	−25.87, −9.12	+7.63, +24.38	39.03	15.88
Reader 2	+8.66	−32.88, +50.20	−53.65, −12.10	+29.43, +70.97	20.74	42.72
Intensity variability						
Reader 1	+21.71	−215.64, +259.06	−334.33, −96.96	+140.38, +377.75	77.00	227.99
Reader 2	+39.25	−349.13, +427.63	−543.34, −154.93	+233.43, +621.83	100.00	374.50
Size zone variability						
Reader 1	−1.88	−57.65, +53.89	−85.53, −29.76	+26.00, +81.78	57.17	52.76
Reader 2	+4.26	−64.77, +73.29	−99.29, −30.25	+38.77, +107.81	64.71	65.68

natural measurement variation in the therapy response setting.

Assessment of the repeatability of MR imaging texture parameters is highly relevant to rectal cancer. MR imaging plays a key role in the treatment of these patients (18). The additional information from texture analysis (3) may contribute to a more personalized management approach by providing prognostic and predictive information. Global first-order statistic-based (mean, standard deviation, skewness, kurtosis, entropy, and energy) and model-based metrics (fractal dimension, lacunarity) have been used most commonly to date because of their availability through commercial and open-access research

software platforms. Initial studies have suggested that global and local-regional features extracted from the primary tumor may predict for nodal metastases (19) and overall survival at CT (20), as well as treatment response at MR imaging (7). De Cecco et al (7) found that T2-weighted filtered global statistical texture parameters differed between responders and nonresponders in an exploratory study of 15 patients with rectal cancer; baseline kurtosis with a medium texture scale was lower in the complete responder group. Complete responders also demonstrated greater changes in kurtosis in the middle of treatment than did nonresponders. A further retrospective exploratory dynamic contrast-enhanced MR imaging–based study of 10 patients with 26 colorectal cancer liver metastases treated with bevacizumab and oxaliplatin–5 fluorouracil–leucovorin that used global statistical and fractal method texture analysis showed that pretreatment low-box dimension is associated with a better tumor response (9).

Nevertheless, there were limitations to our study. We chose to analyze T2-weighted images acquired by using a turbo spin-echo sequence specifically because these sequences are performed universally in clinical practice and have been shown to provide predictive and prognostic information in rectal cancer (18,21,22); however, our findings may not be generalizable to other MR imaging sequences that may be performed to assess rectal cancer (including T1-weighted, diffusion-weighted, and dynamic contrast-enhanced MR imaging). It has been shown that texture analysis can be affected by the MR imaging acquisition and postprocessing methods (23–25); spatial resolution may be the most important factor for MR imaging (23,24). However, one group of researchers (24) has suggested that if the spatial resolution is sufficiently high, variations in repetition time, echo time, and number of averages have little effect; indeed, texture features derived from the co-occurrence matrix were the most robust in their study (23). Other groups have suggested that signal-to-noise ratio, field strength

(1.5 vs 3.0 T), and intensity normalization algorithms (25) influence texture analysis, highlighting the importance of consistent methods. In our study, the MR imaging sequences were performed as part of a prospective research protocol and were therefore subject to standardization and quality control. Indeed, the same MR imaging unit and operator were used for each patient. Therefore, our results reflect a "best-case scenario" that may not necessarily be directly translatable to different MR imaging units with different protocols. However, the knowledge of the degree of variation in general remains relevant to future clinical practice. Our patient numbers were also small, reflecting the challenges of obtaining consent for a second acquisition that will not alter treatment and that is not perceived as having a direct clinical benefit.

In conclusion, few studies to date have assessed the repeatability of image texture parameters. We found that the day-to-day fluctuations in MR imaging texture parameters were lower for global parameters than for local-regional parameters. In particular, first-order histogram (mean, median, entropy) and model-based fractal parameters (fractal dimension mean, fractal dimension standard deviation) consistently demonstrated good repeatability, with narrow limits of agreement and a wCV of less than 10%, which is acceptable for clinical practice.

References

1. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366(10):883–892.

2. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.

3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278(2):563–577.

4. Kuo MD, Jamshidi N. Behind the numbers: decoding molecular phenotypes with radiogenomics—guiding principles and technical considerations. Radiology 2014;270(2):320–325.

5. Davnall F, Yip CS, Ljungqvist G, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging 2012;3(6):573–589.

6. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging 2013;40(1):133–140.

7. De Cecco CN, Ganeshan B, Ciolina M, et al. Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. Invest Radiol 2015;50(4):239–245.

8. Parikh J, Selmi M, Charles-Edwards G, et al. Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy. Radiology 2014;272(1):100–112.

9. O'Connor JP, Rose CJ, Jackson A, et al. DCE-MRI biomarkers of tumour heterogeneity predict CRC liver metastasis shrinkage following bevacizumab and FOLFOX-6. Br J Cancer 2011;105(1):139–145.

10. Alic L, van Vliet M, van Dijke CF, Eggermont AM, Veenland JF, Niessen WJ. Heterogeneity in DCE-MRI parametric maps: a biomarker for treatment response? Phys Med Biol 2011;56(6):1601–1616.

11. Ashraf AB, Daye D, Gavenonis S, et al. Identification of intrinsic imaging phenotypes for breast cancer tumors: preliminary associations with gene expression profiles. Radiology 2014;272(2):374–384.

12. Fried DV, Tucker SL, Zhou S, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. Int J Radiat Oncol Biol Phys 2014;90(4):834–842.

13. Balagurunathan Y, Kumar V, Gu Y, et al. Test-retest reproducibility analysis of lung CT image features. J Digit Imaging 2014; 27(6):805–823.

14. Sanghera B, Banerjee D, Khan A, et al. Reproducibility of 2D and 3D fractal analysis techniques for the assessment of spatial heterogeneity of regional blood flow in rectal cancer. Radiology 2012;263(3):865–873.

15. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour ^{18}F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. Eur J Nucl Med Mol Imaging 2013;40(11):1662–1671.

16. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. J Nucl Med 2012;53(5):693–700.

17. Bland JM, Altman DG. Measurement error proportional to the mean. BMJ 1996;313 (7049):106.

18. MERCURY Study Group. Diagnostic accuracy of preoperative magnetic resonance imaging in predicting curative resection of rectal cancer: prospective observational study. BMJ 2006;333(7572):779.

19. Cui C, Cai H, Liu L, Li L, Tian H, Li L. Quantitative analysis and prediction of regional lymph node status in rectal cancer based on computed tomography imaging. Eur Radiol 2011;21(11):2318–2325.

20. Ng F, Ganeshan B, Kozarski R, Miles KA, Goh V. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. Radiology 2013;266(1):177–184.

21. 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):3753–3760.

22. Taylor FG, Quirke P, Heald RJ, et al. Preoperative magnetic resonance imaging assessment of circumferential resection margin predicts disease-free survival and local recurrence: 5-year follow-up results of the MERCURY study. J Clin Oncol 2014;32(1):34–43.

23. Mayerhoefer ME, Szomolanyi P, Jirak D, Materka A, Trattnig S. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study. Med Phys 2009;36(4):1236–1243.

24. Waugh SA, Lerski RA, Bidaut L, Thompson AM. The influence of field strength and different clinical breast MRI protocols on the outcome of texture analysis using foam phantoms. Med Phys 2011;38(9):5058–5066.

25. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging 2004; 22(1):81–91.