Daily quantitative MRI for radiotherapy response monitoring

Kooreman, E.S.

FEASIBILITY AND ACCURACY OF QUANTITATIVE IMAGING ON A 1.5 T MR-LINEAR ACCELERATOR

Ernst S. Kooreman, Petra J. van Houdt, Marlies E. Nowee, Vivian W.J. van Pelt, Rob H.N. Tijssen, Eric S. Paulson, Oliver J. Gurney-Champion, Jihong Wang, Folkert Koetsveld, Laurens D. van Buuren, Leon C. ter Beek, Uulke A. van der Heide

Radiotherapy & Oncology 133, 156-162 (2019)

ABSTRACT

Purpose: Systems for magnetic resonance (MR-) guided radiotherapy enable daily MR imaging of cancer patients during treatment, which is of interest for treatment response monitoring and biomarker discovery using quantitative MRI (qMRI). Here, the performance of a 1.5 T MR-linac regarding qMRI was assessed on phantoms. Additionally, we show the feasibility of qMRI in a prostate cancer patient on this system for the first time.

Materials and methods: Four 1.5 T MR-linac systems from four institutes were included in this study. T_1 and T_2 relaxation times, and apparent diffusion coefficient (ADC) maps, as well as dynamic contrast enhanced (DCE) images were acquired. Bland-Altman statistics were used, and accuracy, repeatability, and reproducibility were determined.

Results: Median accuracy for T_1 ranged over the four systems from 2.7 to 14.3 %, for T_2 from 10.4 to 14.1 %, and for ADC from 1.9 to 2.7 %. For DCE images, the accuracy ranged from 12.8 to 35.8 % for a gadolinium concentration of 0.5 mM and deteriorated for higher concentrations. Median short-term repeatability for T_1 ranged from 0.6 to 5.1 %, for T_2 from 0.4 to 1.2 %, and for ADC from 1.3 to 2.2 %. DCE acquisitions showed a coefficient of variation of 0.1 to 0.6 % in the signal intensity. Long-term repeatability was 1.8 % for T_1, 1.4 % for T_2, 1.7 % for ADC, and 17.9 % for DCE. Reproducibility was 11.2 % for T_1, 2.9 % for T_2, 2.2 % for ADC, and 18.4 % for DCE.

Conclusion: These results indicate that qMRI on the Unity MR-linac is feasible, accurate, and repeatable which is promising for treatment response monitoring and treatment plan adaptation based on daily qMRI.

INTRODUCTION

Biomarkers derived from quantitative Magnetic Resonance Imaging (qMRI) are promising for oncology, where they provide functional information for treatment response monitoring and prediction. Dynamic contrast-enhanced (DCE-) MRI, which measures perfusion and permeability, and diffusion weighted imaging (DWI) MRI have shown to provide valuable information for different types of cancer [27,28]. Moreover, these modalities are increasingly recommended in guidelines for tumor staging and treatment response monitoring [29,30].

The use of qMRI biomarkers in clinical practice is currently limited [12]. Differences between systems, sequences, image reconstruction algorithms, and data processing methods all influence the results of biomarker studies and complicate the comparison of studies from different centers [14,31]. Moreover, studies investigating the use of qMRI biomarkers as surrogate endpoints have been limited to small patient groups. Larger cohort studies relating qMRI biomarkers to clinical outcome are yet to be done.

An additional challenge for treatment response monitoring, is that imaging studies often are performed at different time points [32–35], and evidence about the optimal timing is lacking. To our knowledge, only one study managed to accomplish daily MRI of patients during radiation treatment of brain metastases [36]. Such studies are challenging to perform, both logistically and in terms of patient burden, but are necessary to acquire more information about MRI related changes due to treatment response.

The recently introduced linear accelerators that are integrated with an MRI scanner (MR-linac), create the possibility of daily imaging during treatment with limited increase of patient burden [37]. Treatments commonly take place over the course of weeks, where patients are daily positioned inside the MR-linac. This makes frequent MR imaging for treatment response monitoring and biomarker studies feasible. However, the design of the MRI scanner of an MR-linac is different from conventional diagnostic systems [1,2,16,38,39], which may influence the quality of the qMRI data. Therefore, before the start of treatment response monitoring and biomarker studies on an MR-linac, a thorough performance assessment is needed. Feasibility of qMRI has been shown previously on a hybrid MR-radiation therapy system that uses a combination of Cobalt-60 sources and a 0.35 T MRI, both qualitatively [40] as quantitatively using DWI [17,18]. These kinds of assessments are important in order to evaluate newly introduced MR-linac systems.

As a first step, this study aims to determine the accuracy, repeatability, and reproducibility of T_1 mapping, T_2 mapping, DWI, and DCE-MRI in phantoms on the Unity MR-linac (Elekta AB, Stockholm, Sweden), which is equipped with a modified Philips 1.5

T MRI system (Philips Healthcare, Best, The Netherlands) [1]. We also show the feasibility of qMRI *in vivo* by acquiring quantitative maps in a prostate cancer patient.

METHODS AND MATERIALS

Study Setup

Four qMRI acquisitions were performed: T_1 mapping, T_2 mapping, DWI for apparent diffusion coefficient (ADC) mapping, and DCE-MRI. We used the recommendations established by the quantitative imaging biomarker alliance (QIBA) DWI- and DCE-MRI profiles [41,42] for assessment of the performance of DWI and DCE-MRI data. For T_1 mapping and T_2 mapping we applied similar approaches as explained in detail below.

Phantom measurements were performed on four Unity MR-linac systems across four institutes. These systems are designated here as MR-linac A, B, C, and D. Each institute used their own copy of the phantoms described below, except institutes B and D, which used the same phantom for DWI.

On all systems, accuracy and short-term repeatability was assessed. The measurements were repeated within the scanning sessions for assessment of the short-term repeatability. To assess long-term repeatability, the measurements were repeated on system A after five months. Reproducibility among the four systems was determined as well. All image analysis and curve fitting were done using MATLAB (Release 2017b, MathWorks, Natick, MA).

In a prospective feasibility study, all qMRI acquisitions were performed on one patient with histologically proven prostate cancer. All protocols were approved by the medical ethics committee of The Netherlands Cancer Institute, and written informed consent was obtained.

T_1 mapping

The Eurospin TO5 phantom (Diagnostic Sonar, Livingston, Scotland) was used to evaluate the performance of T_1 mapping (see Figure 1a). Twelve gel samples were chosen with T_1 relaxation times between 329 and 1603 ms at 296 K at 1.5T. The variable flip angle (VFA) method was applied [43], using a spoiled gradient echo sequence with flip angles of 3, 6, 10, 20, and 30°. This method was chosen because it provides a fast way to map a 3D volume and is therefore often used clinically. The remaining acquisition parameters can be found in Table S2.1.

T_1 values were estimated from the mean values of regions of interest (ROIs) with a diameter of 12 mm per tube by a linear least-squares method [44]. Two institutes

switched one tube with a tube of the DCE phantom for baseline T_1 measurements, so this tube was omitted in the analysis for all institutes. The temperature was measured before and after each acquisition in a tube with water that was kept near the phantom during all experiments. The average temperature was used to correct estimated T_1 values to the reference value at 296 K. For each tube, this reference value was provided by the phantom manufacturer at a field strength of 1.5T and a temperature of 296 K.

T_2 mapping

The Eurospin TO5 phantom was used for T_2 mapping as well, but with a different set of gel samples (see Figure 1b). The T_2 values ranged between 49 and 212 ms at 296 K at 1.5 T. For acquisition, an accelerated multi-echo spin echo sequence was used. Acquisition parameters differed slightly for each system and can be found Table S2.2. Average decay curves were constructed for ROIs with a diameter of 12 mm within each tube. The T_2 values were estimated by fitting a mono-exponential decay function with a nonlinear least-squares method. To avoid bias from stimulated echoes, the first echo was discarded for analysis which was achieved by skipping the acquisition of the first echo during scanning [45].

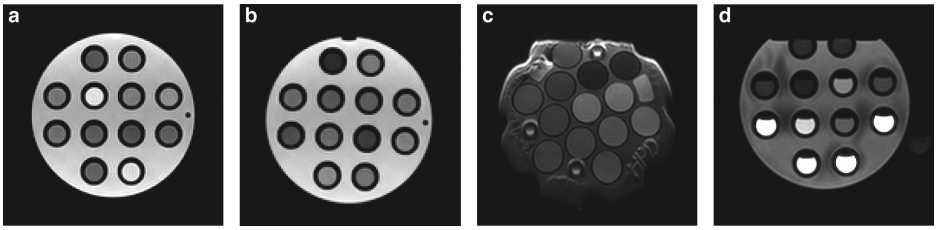

Figure 2.1. MR images of the T_1, T_2, ADC, and DCE phantoms. (a) T_1-weighted VFA image with flip angle of 10 degrees of the Eurospin TO5. (b) T_2-weighted image with a TE of 70 ms. (c) DWI image (b = 0 s/mm^2) of the QIBA phantom at the iso-center. (d) T_1-weighted image of the DCE phantom.

As with the T_1 mapping, the temperature was monitored before and after each experiment, and the average temperature was used to correct estimated T_2 values to a reference value at 296 K.

DWI

A diffusion phantom (High Precision Devices, Inc, Boulder, Colorado) recommended by the QIBA DWI profile was used for DWI measurements (see Figure 1c) [41]. This phantom contains 13 separate vials with aqueous solutions of 0 – 50 % w/w polyvinylpyrrolidone. The vials were surrounded by ice water to ensure measurements at 0 °C. The phantom was placed such that the central tube was in the iso-center of the system. Each institute scanned the QIBA recommended calibration protocol, which uses a spin-echo echo-

planar imaging acquisition. Diffusion weighing was achieved using Stejskal-Tanner diffusion gradients with four b-values: 0, 500, 900, and 2000 s/mm2. Other acquisition parameters can be found in Table S2.3. Additionally, a clinical protocol with larger voxels and b-values of 0, 200, and 800 was scanned on systems A and D (Table S2.7). Apparent diffusion coefficient (ADC) maps were calculated offline by fitting a linear function to the log of the signal decay versus b-values using linear least-squares. Mean values of the tubes were determined in a ROI with a diameter of 13 mm and compared to values provided by the phantom manufacturer.

DCE

The QIBA DCE profile [42] proposes to test three aspects of the acquisition: the accuracy of T_1 estimation, the stability of the signal during acquisition, and the linearity between signal intensity and concentration of the contrast agent. The first aspect was tested with the T_1 mapping of the Eurospin TO5 phantom. For the latter two, a phantom was created consisting of ten tubes with different concentrations of gadolinium between 0 and 9.8 mM (Dotarem, Guerbet, France, T_1 relaxivity 3.9 s^{-1} mM^{-1}) dissolved in a stock solution of water and 0.045 mM manganese chloride. These tubes were inserted in the Eurospin TO5 holder for image acquisition (see Figure 1d).

A spoiled gradient-echo sequence was repeated for 4:39-5:17 minutes with a temporal resolution of 4.1-4.7 s, for a total of 65-75 scans. Additional acquisition parameters can be found in Table S2.4. Both the stability of the signal over all scans, and the relation between signal intensity and concentration was assessed. The signal intensity was measured as the mean of a ROI with a diameter of 5 mm in each tube. To ensure a steady state, the first two dynamics were discarded. The median value of the remaining dynamics was used for analysis.

Although the QIBA DCE profile suggests to assess the linearity of the signal intensities over the tubes with a range of contrast agent concentrations, we converted the signal intensities to concentration values to be able to compare the results between systems [46]. For this, a baseline T_1 value is needed, which was represented by the tube with 0 mM gadolinium. The T_1 value of this tube was determined separately using an inversion recovery (IR) series with inversion times of 30 – 4000 ms on each MR-linac individually. Acquisition parameters of the IR series are given in Table S2.5. The IR method was used because it is regarded the gold standard [47,48], so the influence of possible inaccuracies of the clinical T_1 mapping method was minimized allowing for a better assessment of the spoiled gradient-echo sequence. The calculated concentration values were compared to the known gadolinium concentrations.

Patient data

All quantitative measurements were obtained *in vivo* in a single patient, with similar settings as described above. Details of the sequence parameters are given in Table S2.6. T_2- and ADC maps were calculated on the system. For DCE imaging, a T_1 map was estimated offline based on a VFA series [46]. The Tofts model was then applied to estimate K^{trans} [49], using an arterial input function with parameters derived from an in-house study population of prostate cancer patients [50].

Statistical Analysis

All statistical analysis was done in R (v 3.4.3). Bland-Altman statistics were used to describe the bias and limits of agreement (LoA) of the accuracy and short-term repeatability for T_1, T_2, and ADC. Kendall's Tau (two-sided) was used to identify dependencies of the variation on the mean value, with a significance level of $\alpha = 0.05$. For parameters with significant dependencies, the relative percent ratio instead of the differences was used for the y-axis of the Bland-Altman plots [51,52]. For accuracy, the difference between the measured and the reference values was plotted as function of the reference value.

Additionally, the accuracy was calculated for each individual tube as the absolute percentage:

$$\text{Accuracy} = \frac{|\text{Measured} - \text{Reference}|}{\text{Reference}} \cdot 100\%, \tag{2.1}$$

with the reference values specified by the phantom manufacturers. For short-term repeatability, Bland-Altman plots were produced by plotting the difference between the first and second measurements (short-term) as a function of the mean of these two values.

Additionally, the repeatability for each individual phantom tube was calculated as follows:

$$\text{Repeatability} = \frac{|\text{Measurement 2} - \text{Measurement 1}|}{\text{Mean}(\text{Measurement 1}, \text{Measurement 2})} \cdot 100\%. \tag{2.2}$$

Short-term repeatability was calculated using the repeated measurements of a single session. Long-term repeatability was calculated using Eq. 2.2 with repeat measurements on MR-linac A with five months in between.

Reproducibility (variation across systems) was quantified for T_1, T_2, and ADC mapping with the % coefficient of variation (%CV):

$$\%CV = \frac{SD(\text{system A,B,C,D})}{Mean(\text{system A,B,C,D})} \cdot 100\%, \quad (2.3)$$

where the standard deviation (SD) and mean of the first measurements on each separate system were used.

Additional measures were calculated to enable comparison with previous studies. For T2, the %CV was calculated for both short- and long-term repeatability, using the SD and mean of the repeat measurements. For the ADC maps, the %CV was calculated by using the SD and mean of the ROI in the tube in the iso-center of the systems. This was determined in the first acquisition of each measurement series. Additionally, diffusion images were analyzed according to the QIBA DWI profile [41].

For DCE, the stability was determined as the %CV, calculated with the ROI means and SDs over the 66-73 remaining scans from the five-minute acquisition.

For all acquisitions, only the slice at the center of the phantom was analyzed.

RESULTS

An overview of the accuracy and short-term repeatability is provided in Figure 2.2. The bias for the accuracy of T_1 (Figure 2.2a) was found to be 11 ms with LoA of ± 238 ms. For the T_1 short-term repeatability (Figure 2.2b), the bias was -6 ms, and the LoA were ± 63 ms. The accuracy, short-term repeatability, long-term repeatability, and reproducibility as calculated according to Eqs. 2.1 – 2.3 are presented in Table 2.1. Except for MR-linac B, the short-term repeatability was found to be lower than the accuracy.

The variation in the accuracy of T_2 (Figure 2.2c) showed a dependence on the T_2 value (τ = .69, p < .001). This was also found for the T_2 repeatability (τ = .44, p < .001) (Figure 2.2d). The bias and LoA, for T_2 were -11 ± 6 and 0 ± 2 % for accuracy and short-term repeatability, respectively. The individual values for each system (Table 2.1) are comparable among systems for both accuracy and short-term repeatability, although the short-term repeatability is much lower.

The variation in ADC repeatability (Figure 2.2f) was found to depend on the measured ADC value (τ = .20, p = .04), so the ratio is shown. Bias and LoA for the accuracy (Figure 2.2e) were 0.007×10^{-3} mm²/s, and ± 0.027×10^{-3} mm²/s. For the short-term repeatability of ADC (Figure 2.2f), the bias was 0 and the LoA ± 9 %. Individual values for the systems (Table 2.1) are similar for both accuracy and short-term repeatability, which in turn are comparable to each other.

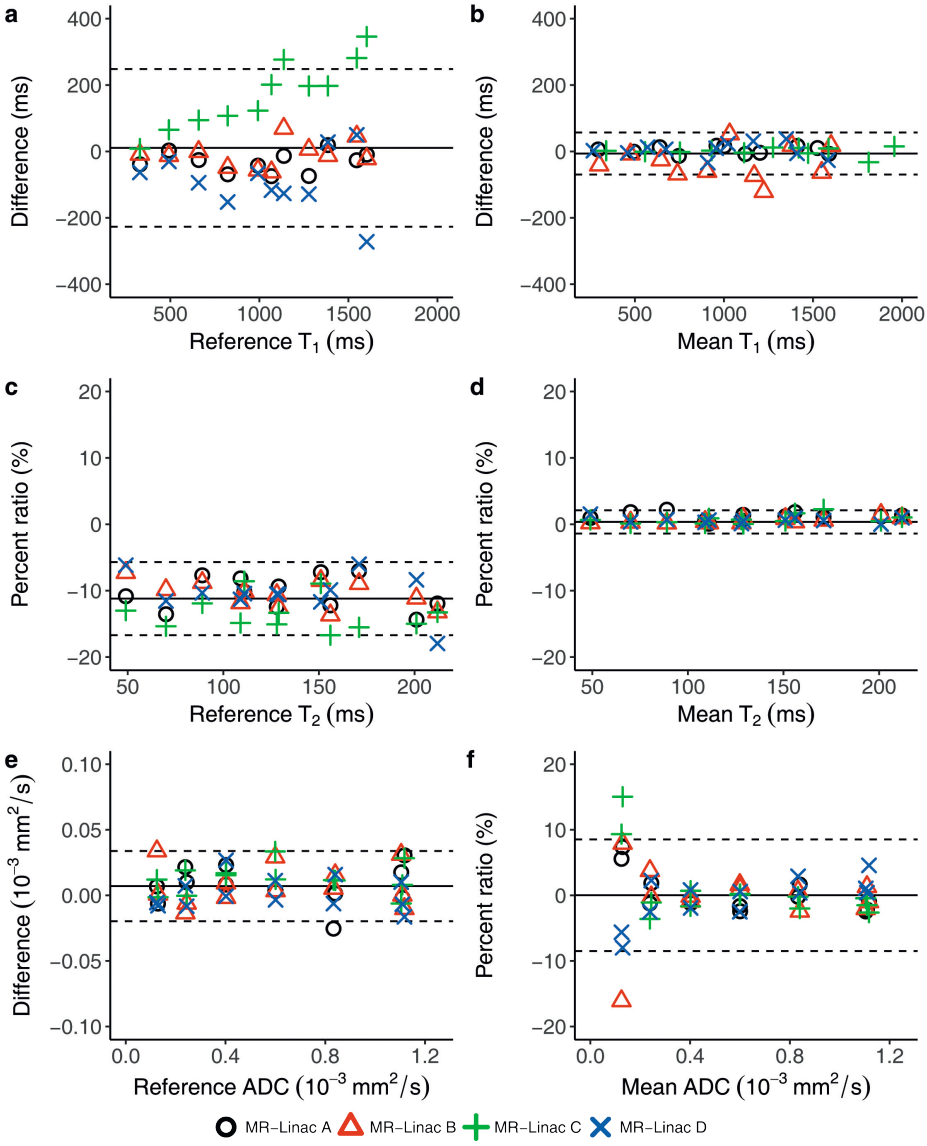

Figure 2.2. Bland-Altman plots showing the accuracy and repeatability of T_1, T_2, and ADC mapping on the MR-linacs. The repeatability shown here (b,d,f) is the short-term repeatability, where two measurements were done consecutively. Each marker corresponds to a measurement in a single tube ROI. The solid line corresponds to the bias, and the dotted lines correspond to the LoA. Note that the y-axes of a, b, and e represent the qMRI units, while c, d, and f represent a percent ratio.

The accuracy in the tube at the iso-center, which is also a measure described in the QIBA DWI profile, was found to be 0.2 %, 0.0 %, 0.7 %, and 0.6 % for MR-linac A, B, C, and D, respectively. The %CV based on the ROI mean and SD in the center tube were found to be 5 % for MR-linac A, 9 % for MR-linac B, 7 % for MR-linac C, and 7 % for MR-linac D. A more complete set of the QIBA DWI profile requirement measures can be found in Table S2.8, as well as results from the clinical protocol with larger voxels.

The DCE stability measurements produced a median %CV, which represents the variation in the signal intensity over the five-minute acquisition, of 0.6 (range: 0.2 – 2.0) % over the 10 tubes in MR-linac A. In MR-linac B the median %CV was 0.1 (range: 0.0 – 1.8) %, in MR-linac C 0.1 (range: 0.0 – 2.5) %, and in MR-linac D 0.6 (range: 0.2 – 2.9) %. Figure 2.3 shows an increasing deviation from the reference value with an increase in concentration. For concentrations of 0.5 mM, the median accuracy was 23.5 (range: 14.8 – 35.5) %. For higher concentrations, this increased to a median accuracy of 62.0 (range: 47.6 – 71.2) % for 9.8 mM. Long-term repeatability was found to be 17.9 (median, range: 1.0 – 37.9) %, and reproducibility (%CV) was 16.7 (median, range: 8.0 – 28.3) %, which is high compared to the other modalities (Table 2.1).

Table 2.1. Accuracy and repeatability of T_1, T_2, and ADC mapping. For each system and qMRI parameter, the median (range) of the measured phantom tubes is given.

	T_1 (%)	**T_2 (%)**		**ADC (%)**
Accuracy				
MR-linac A	4.0 (0.6 – 11.8)	10.5 (7.0 – 14.4)		2.7 (0.2 – 9.0)
MR-linac B	2.7 (0.1 – 6.1)	10.4 (7.3 – 13.7)		1.9 (0.0 – 27.1)
MR-linac C	14.3 (2.6 – 24.4)	14.1 (8.6 – 16.7)		2.0 (0.1 – 9.6)
MR-linac D	10.9 (2.0 – 19.2)	10.5 (6.0 – 18.0)		1.9 (0.2 – 6.6)
Short-term repeatability			**%CV**	
MR-linac A	1.2 (0.2 – 2.1)	1.2 (0.1 – 2.2)	0.8 (0.1 – 1.6)	1.7 (0.2 – 7.1)
MR-linac B	5.1 (1.1 – 13.8)	0.4 (0.2 – 1.5)	0.3 (0.1 – 1.0)	1.3 (0.2 – 17.5)
MR-linac C	0.6 (0.1 – 1.8)	0.6 (0.1 – 2.2)	0.4 (0.0 – 1.6)	1.5 (0.1 – 14.0)
MR-linac D	1.7 (0.4 – 3.7)	0.6 (0.0 – 1.5)	0.4 (0.0 – 1.1)	2.2 (0.4 – 8.3)
Long-term repeatability				
MR-Linac A	1.8 (0.4 – 5.6)	1.4 (0.8 – 2.9)	1.0 (0.6 – 2.1)	1.7 (0.0 – 6.4)
Reproducibility (%CV)				
All systems	11.2 (6.6 – 15.8)	2.9 (0.9 – 4.7)		2.2 (0.6 – 12.0)

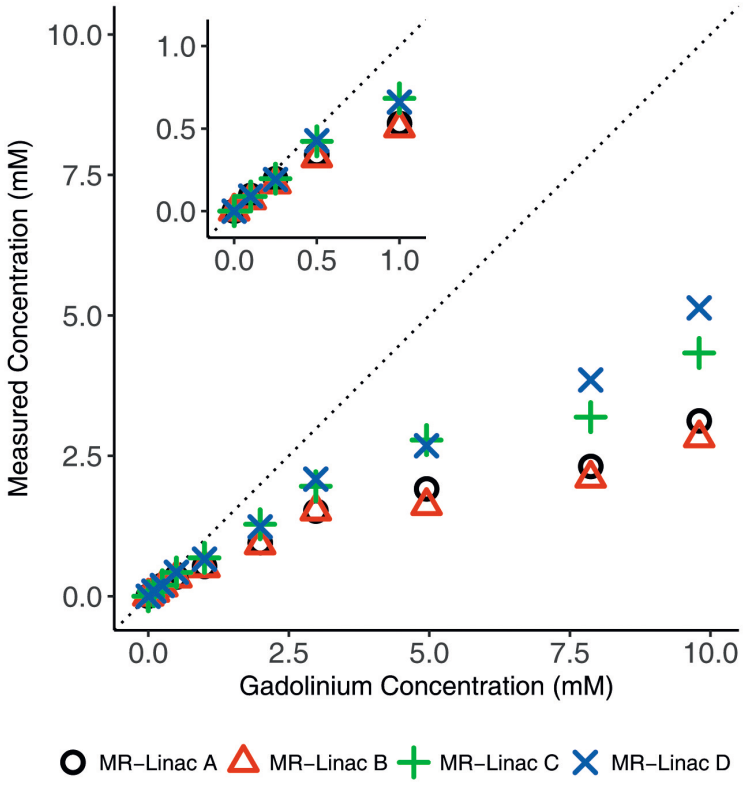

Figure 2.3. Measured gadolinium concentration plotted against the reference value. The dotted line represents the identity line (y = x).

Figure 2.4 shows the quantitative maps from one prostate cancer patient (62 years, initial PSA 37 µg/L, Gleason score 4+5), acquired before the start of radiation treatment. The tumor is clearly visible on the T_2, ADC, and K^{trans} maps (Figure 2.4 b-d), and the images indicate good quality with minimal distortions and no obvious artefacts.

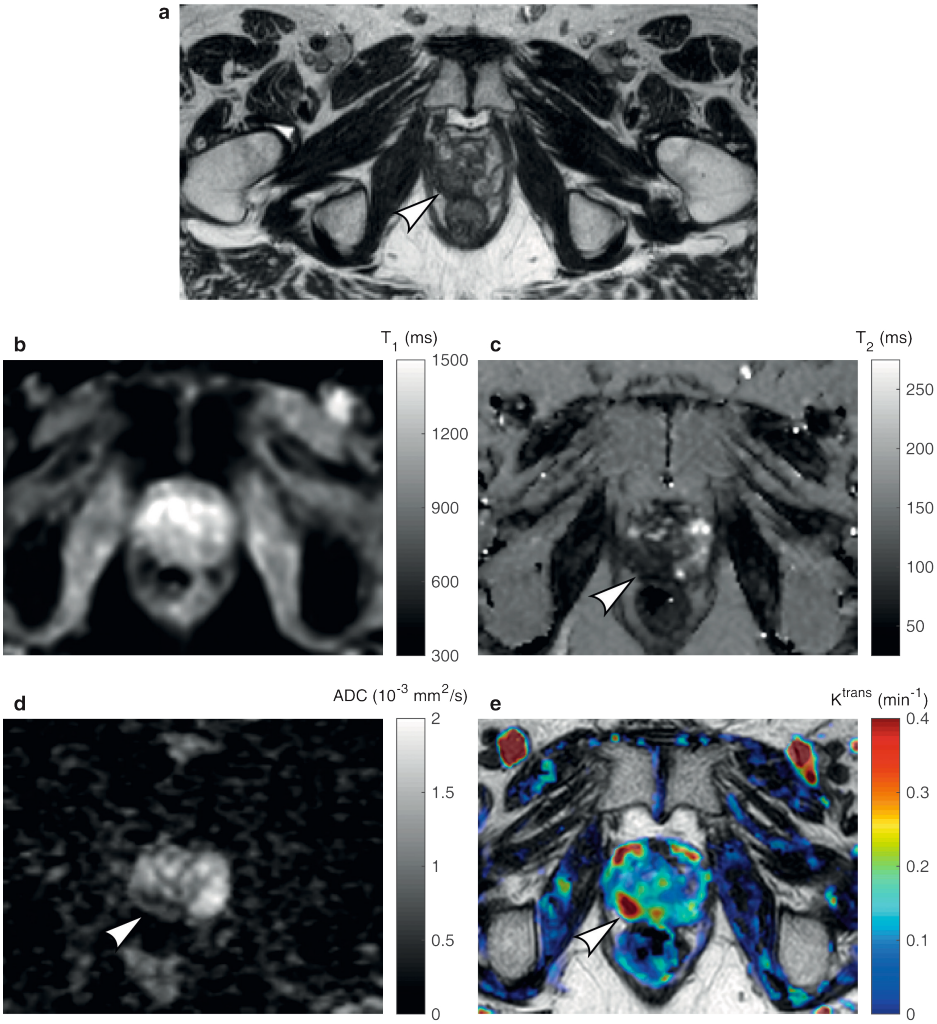

Figure 2.4. Quantitative maps of a patient with prostate cancer. (a) T_2-weighted anatomical image, (b) T_1 map, (c) T_2 map, (d) ADC map, (e) K^{trans} map. The white arrows indicate the location of the tumor, visible on the T_2 map, ADC map, and K^{trans} map.

DISCUSSION

MR-linac systems may enable daily qMRI acquisitions for treatment response monitoring, prediction, and biomarker discovery during radiotherapy. In this study, we determined the accuracy, repeatability, and reproducibility of quantitative T_1 mapping, T_2 mapping, ADC mapping, and DCE-MRI on the Unity MR-linac. Additionally, we are the first to demonstrate feasibility of these quantitative acquisitions on the Unity MR-linac in a patient with prostate cancer.

The Bland-Altman plots provide an overview of the four MR-linacs, which means that the reported biases and LoAs summarize the group, and do not represent individual systems. For T_1 and T_2 they however show clearly that repeated measurements on a single system (Figure 2.2b,d) show less variation than measurements between systems (Figure 2.2a,c). The former is interesting for single-center studies, where one system is used to assess variations in individual patients over time, and the latter is interesting for multi-center studies.

The accuracy of the T_1 VFA series (Figure 2.2a) shows great variation between systems, especially for higher T_1 values. This is likely the result of using the VFA method, which is known to depend on several parameters, e.g. spoiling, and is generally known to overestimate T_1 values [47,48]. The variation between systems underwrites that careful validation and optimization is needed if the VFA were to be used in a multi-center study.

For T_2 accuracy, a clear negative bias was found over the entire range of T_2 values, which was also found previously in studies using a multi-echo approach [53–55]. As this bias is over the entire range of T_2 values, this should be of little influence to detect differences in tissues. The LoA indicate a variation of 12 %, which corresponds to 6-25 ms over the range of assessed T_2 values. Therefore, differences between for instance a prostate tumor (80 ms) and healthy prostate tissue (150 ms) [56,57] should be very well detectable. The zero bias and narrow LoA of the repeatability indicate that small changes in T_2 due to radiation treatment should be detectable. Short- and long-term repeatability %CV are comparable to the results of diagnostic systems [57].

For ADC mapping, the QIBA DWI profile presents threshold values for measurements in the iso-center that represent requirements for systems to meet the profile claims about confidence intervals for ADC measurements in patients [41]. For the accuracy, all systems passed the requirement of ≤ 3.6 %, as the highest value found was 0.7 %, indicating that ADC measurements in a ROI are accurate. This center tube accuracy also compares to previously found values in diagnostic systems [58,59]. On the other hand, none of the systems met the requirement for the %CV, as this was above the recommended 2 % for all systems. This was also found previously on a 1.5 T system [60]. One reason for the increased ROI %CV values is the reduced signal to noise ratio of the available 8-channel body array [16] compared to a head coil as recommended by the QIBA. The %CV is determined using the SD of the ROI in the central tube and is therefore closely related to the SNR. Indeed, evaluation of the clinical acquisition sequence with larger voxels and lower maximum b-value showed that the %CV requirements were met (Table S2.8). For the other tubes, the accuracy deteriorated up to 27.1 % in MR-linac B for the vial with the lowest ADC (0.125×10^{-3}) mm^2/s). The repeatability also worsened at these low ADC values. These deviations can partly be explained by lower SNR due

to the receiver coil, but possibly also by gradient nonlinearities which influence ADC values measured away from the iso-center [58]. This influence, as well as system based geometric distortions of the Unity MR-linac should be quantified, especially if the goal for imaging is treatment planning or dose painting [61]. Short- and long-term repeatability of ADC mapping in the iso-center are comparable to previously published results [58].

For DCE, good stability was found, with a %CV in signal intensity below 2 % over all tubes over the course of five minutes. The gadolinium concentrations were determined inaccurately, where concentrations of 0.5 mM could be estimated within a range of ~30 %. For higher gadolinium concentrations, the systems all show a severe underestimation. This should not be a problem for low-perfused organs like the prostate but might result in an under-estimation for well-perfused tissues. Although the errors in repeatability and reproducibility are relatively high, the patient K^{trans} image shows the added value on the single patient level, as the tumor is clearly visible (Figure 2.4e).

In conclusion, we assessed the performance of the Unity MR-linac for a range of quantitative MR sequences and showed the feasibility of qMRI in a single patient. The accuracy and repeatability for T_1 and T_2 are similar to literature values from diagnostic systems. ADC mapping is also accurate although larger voxels might be advisable to increase the SNR. DCE acquisitions are decreasingly accurate for increasing contrast agent concentration but are stable and valuable for individual patients. This indicates that the Unity MR-linac performs similar to diagnostic MRI systems and can be used for treatment response monitoring and biomarker discovery studies.

SUPPLEMENTARY MATERIAL

Table S2.1: T_1-map sequences

System	TR (ms)	TE (ms)	Flip Angle (°)	FOV (mm^3)	Parallel Imaging	Acquired Voxel Size (Reconstructed) (mm^3)	Acquisition Time (1 FA) (s)
MR-Linac A	20	4	3, 6, 10 ,20, 30	256 x 256 x 20	No	2 x 2 x 4 (2 x 2 x 2)	11
MR-Linac B	20	4	3, 6, 10 ,20, 30	256 x 256 x 20	No	2 x 2 x 5 (2 x 2 x 2)	11
MR-Linac C	20	2.4	3, 6, 10 ,20, 30	260 x 260 x 150	Yes, 2x	2 x 2 x 10 (1.8 x 1.8 x 5)	19
MR-Linac D	20	4	3, 6, 10, 20, 30	256 x 256 x 20	No	2 x 2 x 4 (2 x 2 x 2)	11

Table S2.2: T_2-map sequences

System	TR (ms)	TE (number of echoes) (ms)	Δ TE	FOV (mm^3)	Parallel Imaging	Acquired Voxel Size (Reconstructed) (mm^3)	Acquisition Time (mm:ss)
MR-Linac A	2693	20 – 190 (18)	10	256 x 256 x 2	Yes, 2x	2 x 2 x 2 (2 x 2 x 2)	03:33[a]
MR-Linac B	2693	20 – 190 (18)	10	256 x 256 x 2	Yes, 2x	2 x 2 x 2 (2 x 2 x 2)	05:47
MR-Linac C	5000	24 – 252 (20)	12	256 x 256 x 2	Yes, 2x	2 x 2 x 2 (2 x 2 x 2)	05:30
MR-Linac D	3439	32 – 192 (10)	16	256 x 256 x 90	Yes, 2x	2.3 x 2.3 x 5 (1.6 x 1.6 x 5)	06:15

[a]A k-t undersampling scheme was used for acceleration ([57])

Table S2.3: DWI sequences according to the QIBA profile [41].

System	TR (ms)	TE (ms)	b-values (s/mm^2)	FOV (mm^3)	Parallel Imaging	Half scan	Acquired Voxel Size (Reconstructed) (mm^3)	Acquisition Time (mm:ss)
MR-Linac A	10000	131	0, 500, 900, 2000	220 x 220 x 124	Yes, 2x	0.622	1.7 x 1.7 x 4 (0.8 x 0.8 x 4)	02:00
MR-Linac B	10000	131	0, 500, 900, 2000	220 x 220 x 124	Yes, 2x	0.619	1.7 x 1.7 x 4 (0.8 x 0.8 x 4)	02:00
MR-Linac C	10000	103	0, 500, 900, 2000	220 x 220 x 124	Yes, 2x	0.651	1.7 x 1.8 x 4 (0.9 x 0.9 x 4)	02:00
MR-Linac D	10000	131	0, 500, 900, 2000	220 x 220 x 124	Yes, 2x	0.622	1.7 x 1.7 x 4 (0.8 x 0.8 x 4)	02:00

Table S2.4: DCE sequences

System	TR (ms)	TE (ms)	Flip Angle (°)	FOV (mm^3)	Parallel Imaging	Dynamic scan time (s)	Number of dynamics	Acquired Voxel Size (Reconstructed) (mm^3)	Acquisition Time (mm:ss)
MR-Linac A	5	1.9	30	220 x 320 x 60	Yes, 2.2x	4.1	75	2.2 x 2.2 x 6 (1.25 x 1.25 x 3)	05:05
MR-Linac B	5	2.4	20	260 x 260 x 100	Yes, 2x	4.7	65	2.5 x 2.5 x 4 (1.35 x 1.35 x 4)	05:17
MR-Linac C	4.4	2.1	25	260 x 260 x 150	Yes, 2x	4.1	69	2 x 2 x 10 (1.8 x 1.8 x 5)	04:39
MR-Linac D	5	1.9	30	220 x 320 x 60	Yes, 2.2x	4.1	75	2.2 x 2.2 x 6 (1.25 x 1.25 x 3)	05:05

Table S2.5: Inversion Recovery for DCE baseline

System	TR (ms)	TE (ms)	TSE Factor	Inversion Time (ms)	FOV (mm^3)	Acquired Voxel Size (Reconstructed) (mm^3)	Acquisition Time (mm:ss)[a]
MR-Linac A	8000	10	7	30, 50, 75, 100, 150, 200, 250, 300, 400, 500, 750, 1000, 1250, 1500, 2000, 4000	256 x 256	2 x 2 x 4 (2 x 2 x 4)	02:40
MR-Linac B	8000	10	7	30, 50, 75, 100, 150, 200, 250, 300, 400, 500, 750, 1000, 1250, 1500, 2000, 4000	256 x 256	2 x 2 x 4 (2 x 2 x 4)	02:40
MR-Linac C	5000	7.8	None	50, 75, 100, 150, 200, 300, 400, 500, 750, 1000, 1250, 1500, 2000, 4000	250 x 250	2 x 2 x 4 (1 x 1 x 4)	10:30
MR-Linac D	8000	10	7	30, 50, 75, 100, 150, 200, 250, 300, 400, 500, 750, 1000, 1250, 1500, 2000, 4000	256 x 256	2 x 2 x 4 (2 x 2 x 4)	02:40

[a]Acquisition times for one inversion time.

Table S2.6: Patient sequences

Sequence	TR (ms)	TE (ms)	Flip Angle	Parallel Imaging	Half-scan	FOV	Acquired Voxel Size (Reconstructed) (mm^3)	Acquisition Time (mm:ss)
T2w (TSE)[a]	1300	114	90	Yes, 3.5x	0.6	400 x 448 x 250	1.2 x 1.2 x 1.2 (0.5 x 05 x 1.2)	06:25 (NSA = 2)
T1 (FFE)	20	4	3, 6, 10, 20, 30	Yes, 3x	No	220 x 251 x 60	2.6 x 2.7 x 7 (1.6 x 1.6 x 3.5)	00:09 (1 FA)
T2 (TSE)[b]	2000	20-190	90	Yes, 2x	0.613	262 x 262 x 60	1.8 x 1.9 x 3 (1.4 x 1.4 x 3)	02:20
DWI (EPI)[c]	3404	61	90	Yes, 2.3x	0.6	430 x 430 x 60	3 x 3 x 3 (1.5 x 1.5 x 3)	04:05
DCE (FFE)[d]	4	1.9	35	Yes, 2x	No	220 x 251 x 60	2.6 x 2.7 x 7 (1.6 x 1.6 x 3.5)	05:05

[a]TSE factor: 110, startup echoes: 4.
[b]Echo times ranged from 20-190 ms with steps of 10 ms for a total of 18.
[c]b-values (averages): 0 (1), 1 (2), 15 (3), 200 (9), 800 (9), only b = 200 and b = 800 were used for ADC reconstruction.
[d]Dynamic scan time = 2.8 s

We suggest in the discussion that the QIBA requirements for the ADC ROI %CV are not met because the system setup requires a head coil which is not available for the Unity MR-linac. Therefore, we acquired additional data on MR-linac A and D using acquisition parameters that increase the SNR (highest b-value 800, shorter TE, increased voxel size, see Table S2.7). The resulting measures show indeed a reduction in the ROI %CV, meeting the requirement on both systems (Table S2.8). Moreover, the requirements for short-term repeatability are also met on both systems.

Table S2.7: DWI sequence with higher SNR.

System	TR (ms)	TE (ms)	b-values (s/mm^2)	FOV (mm^3)	Parallel Imaging	Half scan	Acquired Voxel Size (Reconstructed) (mm^3)	Acquisition Time (mm:ss)
MR-Linac A & D	8462	60	0, 200, 800	430 x 430 x 240	Yes, 2.3x	0.6	3 x 3 x 4.8 (1.3 x 1.3 x 4.8)	04:39

Table S2.8: Overview of measures calculated according to the QIBA DWI profile [41]. A bold font marks QIBA requirements that are met. Values between parentheses (MR-linac A and D) are derived from the more clinical protocol (Table S2.7).

	QIBA requirement	MR-linac A	MR-linac B[a]	MR-linac C	MR-linac D
ADC ROI bias	≤3.6 %	**0.2** %	**0.0** %	**0.7** %	**0.6** %
ADC ROI %CV	≤2 %	5 (**1**) %	9 %	7 %	7 (**2**) %
Short-term ADC repeatability	RC≤1.5 ×10^{-5} mm^2/s wCV≤0.5 %	RC=3.0 (**1.0**) wCV=1.0 (**0.3**)	RC=2.2 wCV=0.9	RC=3.3 wCV=1.1	RC=3.1 (**0.3**) wCV=1.0 (**0.1**)
Long-term ADC repeatability	RC≤6.5 ×10^{-5} mm^2/s wCV≤2.2 %	**RC=2.5** **wCV=0.8**			
DWI b=0 SNR	≥50±5	**101±11**	25±4	**52±5**	**61±7**
ADC b-value dependence	≤2 %	**0.5** %	**0.7** %	**0.2** %	**0.1** %

[a]Three instead of four intra-session measurements were available from MR-linac B.