Plan Quality Assurance for a 1.5 T MR-linac

Viktoriia Gorobets
Utrecht, The Netherlands

ABSTRACT

Context. Quality assurance in radiotherapy involves a process of verification to ensure that each patient receives treatment as specifically prescribed by their radiation oncologist. Currently, various 3D quality assurance systems that are compatible with MRI technology are in use today, including ArcCHECK-MR and Delta4-MR. Additionally, PTW is introducing the OCTAVIUS 4D MR system. This research presents a basic characterization of the OCTAVIUS 4D MR system and a comparative analysis of three MR-compatible 3D QA devices for IMRT Plan QA: ArcCHECK-MR, Delta4-MR, and OCTAVIUS 4D MR with OCTAVIUS Detector (OD) 1500 MR.

Goal. The primary goal of this research was to evaluate the performance of various 3D arrays for quality assurance in radiotherapy within a 1.5 T MR-linac and to identify any limitations these devices might have for patient-specific quality assurance. Before performing the patient quality assurance, a basic characterization of the OCTAVIUS 4D MR with OD 1500 MR and OD 1600 MR detector plates in a 1.5 T MR-linac was performed. 25 patients' treatment plans across different target sizes and locations were measured and compared using ArcCHECK-MR, Delta4+ MR, and OCTAVIUS 4D MR with OD 1500 MR. A comparison of the standard clinical plans and plans containing monitor units (MU) and position errors was performed. The impact of device geometry on the results was explored by analyzing the gamma percentage difference, mean gamma index, and other metrics provided by the software.

Method. For the evaluation, gamma passrate and mean gamma index were obtained by 3D global gamma comparison analysis with 3%/3mm and 2%/2mm criteria. Nominal thresholds of 95% and 90%, respectively, were applied. For intentionally erroneous plans additional metrics were collected such as dose difference (median, mean) and distance to agreement where possible.

Results. The results of the basic characterization of the OD 1500 MR and OD 1600 MR demonstrated no significant differences in short-term reproducibility (< 0.2%), dose linearity (< 1%), field size dependency (< 0.7% for field sizes larger than 5 cm × 5 cm), dose rate dependency (< 0.8%), dose-per-pulse dependency (< 0.4%) and angular dependency (standard deviation < 0.5%).

In the analysis of plans with deliberately introduced errors, the Delta4+ MR showed higher variation in results, the ArcCHECK-MR showed greater error sensitivity. All systems were able to identify position errors that were out of tolerance level, however, bigger targets were less sensitive to misplacements. The gamma test was not particularly effective at detecting dose errors in any of the systems in which it was used, especially underdose in small target regions or highly modulated beams. The gamma test should be used in combination with a dose assessment to identify these kinds of inconsistencies.

Conclusions. The new MR compatible OCTAVIUS detectors and the OCTAVIUS 4D Phantom MR are suitable for QA of patient treatment plans in a 1.5T MRI-linac and for measurements with the offset.

All devices demonstrated their capability to produce good results for standard clinical plans. However, minor misalignment during the setup procedure could affect the accuracy of the analysis. The limited resolution and density of the detectors can impact the precision of results in plans with small high-dose regions. When investigating plans with introduced errors, gamma comparison analysis alone was revealed to be insufficient for MU error detection and should be combined with a dose assessment. Nevertheless, all devices showed a worsening in results for plans with deliberately introduced errors. The responses of the ArcCHECK-MR and Delta4+ MR were more sensitive but also had larger uncertainty.

1 INTRODUCTION

The field of radiation has seen significant changes in the last ten years. Developments in intensity-modulated radiation therapy (IMRT) for accurate dose delivery, improved daily position verification, and three-dimensional tumor visualization have made it easier to increase local doses while reducing side effects in normal tissues. Continued advancements have resulted in the evolution of cone-beam computed tomography (CT) for lung stereotactic therapies [57], the integration of implanted fiducial markers for image-guided radiotherapy in prostate plans [14], and the application of magnetic resonance imaging (MRI)-guided brachytherapy for the cervix [23, 45].

A significant achievement was reached through the collaboration of the University Medical Center Utrecht, Elekta AB (Stockholm, Sweden), and Philips (Best, The Netherlands), resulting in the development of a linear accelerator (linac) integrated with magnetic resonance imaging (MRI). This system allows for simultaneous irradiation and precise real-time image guidance with soft-tissue contrast for online therapy guidance in oncology [3, 24, 47]. However, the presence of the magnetic field affects secondary charged particles, changing their trajectory perpendicular to their original motion and influencing dose accuracy [46, 61]. This problem has resulted in the development of novel approaches and tools for the commissioning and quality assurance (QA) of MR-linac systems.

Quality assurance in radiotherapy involves a process of verification to ensure that each patient receives their treatment as specifically prescribed by their radiation oncologist [63]. A complete evaluation, which includes comparing measured and calculated dose distributions, must be performed in 3D space. Currently, various 3D quality assurance systems that are compatible with MRI technology are in use today, including ArcCHECK-MR and Delta4-MR. Additionally, PTW is introducing the MR-compatible Octavius 4D system, further expanding the options for MRI-compatible QA in radiotherapy.

Previously, the conventional versions of Delta4 [9, 49, 50], ArcCHECK [5, 19, 27, 52], and OCTAVIUS 4D, along with the OCTAVIUS Detector 1500 array, have been evaluated without a magnetic field [35, 56, 59, 60].

ArcCHECK-MR was the first MR-compatible QA system and has undergone basic characterization in the presence of up to 1.5 T magnetic field [20, 29]. The study showed a dependence on field size with a maximum discrepancy of 2.7% for fields larger than 5 × 5 cm^2. Following this, the Delta4+ MR was released and characterized in magnetic fields of 0.35 T [15] and 1.5 T [13]. Observations indicated an increase in measurement deviations as field size decreases, with discrepancies up to 3% for fields smaller than 5x5 cm. The performance characteristics of the OCTAVIUS Detector 1500 MR have not yet been fully evaluated for all fundamental performance characteristics except for the angular sensitivity with the older phantom [37]. Hence, the results of these basic tests using the new OCTAVIUS 4D MR are reported in this work.

Numerous studies have examined the comparison of conventional devices [22, 30, 51, 55, 62], and there have been investigations into the sensitivity of these devices to various errors that can arise during calibration, issues with a linac, or phantom-related problems [2, 6, 17, 54]. It has been demonstrated that variations in detector resolution and interpolation techniques can lead to discrepancies in the results.

Desai et al. [15] conducted a study on a comparison between MR-compatible Delta4-MR and ArcCHECK-MR. Investigations showed that under the more stringent 2%/2mm global gamma requirements, the differences between the devices become more apparent. The Delta4 Phantom+ MR has a strong correlation between its gamma pass rates and a variety of plan complexity parameters. Meanwhile, the ArcCHECK-MR's performance at the 2%/2mm global gamma level appears to be fairly independent of the complexity of the radiation treatment plan.

PTW is introducing the OCTAVIUS 4D MR system, which is specifically designed to rotate, ensuring a constant and optimal alignment between detectors and each beam segment direction. This characteristic is especially important in treatment settings that use MRI technology. The fully MR-compatible systems are now available for testing and implementation in MR-Linacs.

While drafting this report, the OCTAVIUS 4D MR is not yet available on the market, but a prototype is accessible for evaluation. Consequently, this paper presents a comparative analysis of three MR-compatible 3D QA devices: ArcCHECK-MR, Delta4-MR, and OCTAVIUS Detector 1500 MR with the OCTAVIUS 4D Phantom MR. Analysis was made using the global gamma comparison analysis with 3%/3mm and 2%/2mm criteria for 25 patient treatment plans with a range of target sizes (rectum, pancreas/liver, lungs, prostate, and lymph nodes). Additionally, we introduce MU errors and isocenter position errors to these plans, simulating MR-linac issues and calibration inaccuracies.

2 BACKGROUND

2.1 Geometry comparison

Understanding how the unique characteristics of these devices, such as geometry, detector type, resolution (Table 1), and software algorithms, impact verification outcomes and quality of IMTR is crucial for effective QA.

Systems with diodes instead of ionization chambers benefit from smaller detectors, which enhance resolution and allow for finer measurements in locations with varying dose distributions. The tiny size of these detectors also allows for a closer arrangement, which improves the detail recorded in measurements [8]. However, diodes are prone to radiation-induced damage, and their results can be influenced by changes in dose rate, dose per pulse at the measuring location, and the energy spectrum of photon radiation [16, 26]. On the other hand, the size of the detection volume for ionization chambers cannot be less than the minimum required to produce a sufficient signal [43, 56]. Moreover, the readings from an ion chamber are affected by the averaging of signals across its sensitive volume and the transportation of secondary electrons through its walls, leading to the phenomenon known as the volume averaging effect [18, 25].

The OCTAVIUS Detector 1500 2D array has all its measurement points within a single 2D plane. The ArcCHECK® arranges its diode detectors in a helical pattern around the exterior of a cylindrical phantom. The Delta4 has diode detectors in two orthogonal planes that form a cross-like structure when viewed in a transverse section of its cylindrical phantom. This means that when evaluations are based only on specific measurement points, there are volumes that remain unmeasured. Consequently, there is a risk that the target or organ at risk could be situated in these unmeasured areas, potentially leading to an unnoticed overdose that exceeds the acceptable tolerance levels. None of the dosimetry systems listed in table 1 are capable of capturing the full 3D dose distribution with high resolution [8].

	OD 1500 MR	**ArcCHECK-MR**	**Delta4+ MR**
Detector Type	Ion. chambers	SunPoint® Diode	p-Si diodes
Detectors Arrangement	Plane-parallel	Helical	Orthogonal planes
No. of Detectors	1405	1386	1069
Active Volume (area)	0.06cm^3	$8 \times 8\,mm^2$	$1 \times 0.05\,mm^2$
Detector Spacing	7.1 mm c-to-c	1 cm	5 mm - center
Max. Field Size	27 x 27 cm^2	Dia./Len.: 21 cm	20 x 20 cm^2

Table 1: Main characteristics of three MR-compatible QA devices.

2.2 3D dose reconstruction algorithm

To approximate the delivered dose in complete 3D, users must rely on calculation algorithms integrated into the software of these systems. These algorithms enable the reconstruction of the full 3D dose distribution inside the phantom and allow for the calculation of doses at detector locations or in space, which are then compared to the planned dose data. This approach offers a clear and straightforward clinical interpretation and serves as an effective tool for IMRT QA measurements [8].

The dose reconstruction process within the entire volume of the cylindrical ArcCHECK algorithm synchronizes the planned dose with measurements from the ArcCHECK device to create a 3D dose distribution within a phantom. It then adjusts the patient's dose calculated by the Treatment Planning System (TPS) based on the ratio of this reconstructed dose to the TPS-calculated dose. The ArcCHECK software SNC Patient was initially characterized [64] and since 2011, the manufacturer has improved both the software and hardware. It now includes angle dependency correction with a virtual inclinometer, field size dependence correction, inhomogeneity correction, and big field measurement merging [11]. The SNC Patient software enhances the accuracy of gamma calculations (described later) by performing an automatic interpolation of the calculated dose to a 1 mm grid. This approach is advised by TG-218 [36] to minimize gamma calculation errors which are dependent on the local dose gradient, the distance between the dose points being evaluated, and the DTA criterion. The guidelines note that when the spacing between dose points is on par with the DTA criterion, gamma calculation errors can be significant in regions where the dose gradient is steep. It's generally recommended that the spacing is at most one-third of the DTA, which is the distance criteria used in gamma calculations. A 1 mm grid spacing offers a practical compromise between computational efficiency and the precision of dose calculations, while also maintaining artifacts at an acceptable level [31].

In the Delta4 ScandiDos software, 3D dose reconstruction is conducted on a per-beam basis using the PDDs from treatment planning system dose. For each beam, the software recalculates the dose distribution in three-dimensional space by scaling it along a series of ray lines. These lines extend from the radiation source to each detector within the Delta4 device. The software assumes that the dose along each ray line linearly correlates with the ratio of the measured dose to the calculated dose at the detector. This process is repeated for all beams, and their respective results are summed to form the overall measured 3D dose distribution. In cases where beams intersect both detector planes of the device, the software uses only one of these planes for the dose-rescaling procedure [19, 50]. The algorithm limits the search to a distance twice as big as the acceptance criteria of the spatial deviation.

The algorithms in the OCTAVIUS software VeriSoft® (PTW) [4, 7] for dose reconstruction require percentage depth dose (PDD) data in water for each beam energy used in the measurements. Using the appropriate PDD for the effective field size, the software reconstructs the dose plane along the beam path for each gantry angle. The PDD data in water is then converted to PDD in polystyrene using a relative electron density. The program combines all measured and predicted dose locations from all gantry angles to generate the phantom's 3D dose distribution. These dose values are then linearly interpolated to create a 3D dose grid, often with a default resolution of 2.5 mm [42]. This stand-alone QA tool is intended to measure dose distributions independently, without relying on the dose grid computed by the TPS. The TPS data are required only for the comparison with measurements [4].

These systems function as "black boxes" to the end-users, meaning the users are not able to change the specific methods used, such as the interpolation technique, the particular DTA (distance to agreement) search algorithm employed, or the total number of voxels evaluated for a given threshold. Several studies have been conducted to compare different devices, taking into account their varying geometries and software algorithms [21, 22, 51, 53, 62].

2.3 Comparison analysis

Although the software mentioned earlier has a variety of metrics to compare measured and calculated doses, they all share the gamma comparison analysis as a common metric.

The gamma index evaluation has become a standard method for comparing measured dose distribution against the dose distribution predicted by commercial treatment planning systems (TPS). It combines dose difference $DD(\mathbf{r}_e, \mathbf{r}_r)$ (Equations 1) and distance to agreement $DTA(\vec{\mathbf{r}_e}, \vec{\mathbf{r}_r})$ (Equations 2) to calculate a dimensionless metric for each point in the evaluated distribution. $\mathbf{r}_e$ and $\mathbf{r}_r$ represent specific points derived from the evaluated dose distribution and the reference distribution, respectively [32], [28].

The dose difference, $\delta(\vec{\mathbf{r}_e}, \vec{\mathbf{r}_r})$, is calculated as

$$\delta(\vec{\mathbf{r}_e}, \vec{\mathbf{r}_r}) = D(\vec{\mathbf{r}_e}) - D(\vec{\mathbf{r}_r}), \tag{1}$$

and the spatial distance as

$$\mathbf{r}(\vec{\mathbf{r}_e}, \vec{\mathbf{r}_r}) = |\vec{\mathbf{r}_e} - \vec{\mathbf{r}_r}|. \tag{2}$$

The generalized Γ function is defined as

$$\Gamma(\vec{\mathbf{r}_e}, \vec{\mathbf{r}_r}) = \sqrt{\frac{\delta^2(\vec{\mathbf{r}_e}, \vec{\mathbf{r}_r})}{\Delta D^2} + \frac{|\vec{\mathbf{r}_e} - \vec{\mathbf{r}_r}|^2}{\Delta d^2}}, \tag{3}$$

where $\triangle D$ and $\triangle d$ are the acceptance criteria for DD and DTA, respectively.

The gamma index is defined as

$$\gamma(\vec{\mathbf{r}_r}) = \min\{\Gamma(\vec{\mathbf{r}_e}, \vec{\mathbf{r}_r})\} \forall (\vec{\mathbf{r}_e}) \tag{4}$$

If the $\gamma \leq 1$, then the evaluated distribution at that specific point is within tolerance. The percentage of points within any given structure that is accepted based on this criterion is evaluated by the gamma passing rate (GPR) [58].

While gamma analysis is a common tool in clinical settings, some studies have expressed doubts about its effectiveness in detecting planning or machine errors [30, 39, 54, 64]. It has been shown that even when deliberate errors are introduced, the gamma analysis pass rate can exceed 90%. Despite its limitations, the gamma metric remains the gold standard in patient-specific quality assurance (PSQA), providing quick analysis in busy clinics and influencing decision-making. Therefore, understanding the impact of gamma analysis in different technical setups is important [42].

In our investigation, the software for each device includes a gamma analysis tool. This tool gives the results in terms of GPR, mean gamma index, maximum gamma index, and maximum DD. VeriSoft software has a 3D gamma calculation algorithm that allows for planar comparisons on a plane-by-plane basis, while also considering adjacent planes. Its primary advantage lies in its capability to perform 3D volumetric analysis. In this approach, traditional two-dimensional pixels are substituted with voxels, facilitating the assessment of the entire volumetric dose distribution within the phantom. Several studies have explored the comparison between 3D and 3D volumetric analysis [42, 48] in the VeriSoft software. These studies indicate that the gamma pass rate tends to be higher

in volumetric analysis due to the significant greater set of information.

SNC Patient software provides gamma comparison analysis in 'relative' and 'absolute' dose modes. Additionally, it provides other metrics like the DTA, available in both modes or the dose difference (DD) between calculated and measured datasets at the normalization point, which is typically the point of maximum dose. ScandiDos provides a 3D gamma pass rate, along with DTA and DD metrics, which also show the median dose difference.

3 MATERIALS AND METHODS

3.1 Basic dosimetric characteristics of the OCTAVIUS 4D MR system

The fundamental characteristics of the detector response for the OD 1500 MR and OD 1600 MR were investigated when used in the OCTAVIUS 4D Phantom MR. Characteristics examined included short-term reproducibility, dose linearity, field size dependency, MU rate dependency, dose-per-pulse dependency, and angular dependency. The evaluation of OD 1500 MR also involved measuring 25 clinical treatment plans across diverse target sizes and anatomical sites. The capability of the OCTAVIUS 4D MR system to measure doses with an offset was tested by evaluating one lung treatment plan with a high-dose region offset. This was done by positioning the OCTAVIUS 4D MR isocenter 5 cm to the left.

The OD 1600 MR was not available for these measurements.

3.2 Treatment planning procedure

For this study, we selected twenty-five patient plans who received IMRT MRgRT treatment at our institute and underwent QA with Delta4-MR. The selection of patients was made to ensure diverse sizes of different anatomical sites and positions. We included five patients for each of the following sites: rectum, pancreas/liver, lungs, prostate, and lymph nodes.

To verify the consistency of results over time, one prostate patient was selected for a repeated QA session using Delta4-MR. This step was taken to compare current findings with those from several years ago, ensuring the older data could be reliably used for this comparison analysis.

All patients plans included in the study underwent QA procedures with the OCTAVIUS 4D MR and ArcCHECK-MR. Treatment plans were made with the Monaco treatment planning system. Before employing the OCTAVIUS 4D MR and OD 1500 MR MR for patient QA, these devices underwent fundamental characterization tests. All measurements were conducted with the 1.5 T Elekta Unity MR-linac.

The phantoms' sensitivity for detecting errors was evaluated by comparing the measured plans against intentionally erroneous plans from the Monaco treatment planning system. These errors were deliberately introduced within the devices' software. MU (monitor unit) errors simulating potential calibration problems were made by modifying the scaling factor. These introduced errors in the TPS plans were varied at levels of +10%, +5%, +3%, -3%, -5%, and -10% compared to the original plans. Position errors were simulated by modifying the isocenter position by 1 mm in the software. The technique used here is a simple simulation of errors to validate the measurement devices.

3.3 SunNuclear ArcCHECK®-MR

The ArcCHECK-MR (Sun Nuclear Corporation, Melbourne, USA) and its accompanying components, including cabling, stand, and accessories, are compatible with magnetic resonance (MR) environments up to 1.5 T. This device has 1386 SunPoint® Diode Detectors, each with an active area of $0.64mm^2$. Structurally, it is designed as a cylinder with an array measuring 21 centimeters in length and diameter. The detectors within the ArcCHECK-MR are spaced 1 cm center-to-center and are capable of measuring the exit and entrance doses during the delivery of radiation. The physical depth of each detector is 2.9 centimeters. The phantom material used in the construction of the ArcCHECK-MR is PMMA (Acrylic). The associated software SNC Patient (version 8.4.1.2) was used for the comparison and visualization. The standard analysis operates in 2D mode, and the software doesn't allow for direct modification to 3D mode. The switch to 3D mode was activated through supplementary files located in the SNC Patient folder.

3.4 Scandidos Delta4 Phantom+ MR

The Delta4 Phantom+ MR (Scandidos, Uppsala, Sweden) consists of two planar detector boards, which are arranged orthogonally in a crossed array formation. These boards are equipped with 1069 p-type silicon diodes, serving as the radiation detection elements. In the high-resolution central area, these diodes are positioned 5 mm apart, while in other regions, they are spaced 1 cm apart. The high-resolution zone of each board has an area of 6 cm x 6 cm, whereas the total area covered by the detector plane is 20 cm x 20 cm. The phantom itself is cylindrical 22 cm in diameter and 40 cm in length. It is made of PMMA. The associated software ScandiDos Delta4 (version 1.00.0240) was used for the comparison and visualization.

3.5 OCTAVIUS Detector 1500 MR array in the cylindrical OCTAVIUS 4D Phantom MR

The OD 1500 MR array (PTW, Freiburg, Germany) consists of 1405 cubic air-filled ion chambers arranged in a plane-parallel configuration, forming a vented chessboard matrix pattern on a 27 x 27 cm^2 grid. Each chamber covers a 4.4 mm x 4.4 mm area with a height of 3 mm, giving it an active volume of approximately 0.06 cm^3. The reference point specified by the manufacturer is located 7.5 mm beneath the surface. The array is equipped with an electrometer capable of accommodating a dose rate from 0.25 Gy/min up to 96 Gy/min. The base plate beneath the ion chambers is constructed from a material nearly equivalent to water (polystyrene), and the detector frame is crafted from glass-reinforced plastic. The associated software VeriSoft (version 8.1.1.0) was used for the comparison and visualization.

3.6 Setup

All devices were calibrated according to the calibration procedure described in the relevant manual. All phantoms were positioned using laser alignment, and the calibtation beams were delivered. Position optimization was employed for a box plan and integrated into the Delta4+ MR measurement results. The plans measured with OCTAVIUS 4D MR underwent automatic alignment per the calculated-measured pair. For ArcCHECK-MR, no further alignments were conducted.

The results for plans with deliberately introduced errors were collected without OCTAVIUS 4D MR automatic alignment since this tool may compromise the device's ability to identify errors and is typically used for calibration beam measurements or 'box' plan in clinical situations.

3.7 Comparison study

The measured data were compared to reference data with a global gamma comparison analysis, using criteria of 3%/3mm and 2%/2mm. The Gamma Passing Rate (GPR) data were collected, with nominal values of 95% for the 3%/3mm criterion and 90% for the 2%/2mm criterion, respectively [36, 41**?**]. A low-dose threshold of 10% was applied for the analysis. The differences between the mean gamma indices for plans with deliberately introduced errors and standard plans were calculated and then averaged for each patient's anatomical group.

3.8 Median and mean dose difference in ArcCHECK-MR and Delta4+ MR

To determine the behavior of different metrics for plans with complicated error cases, such as lymph node plans with small target size, DD (Dose Difference), DTA, and mean DD metrics were used for the Delta4+ MR. In the ArcCHECK-MR software, the additional metric of DD at the normalization point and mean DD were collected.

4 RESULTS

4.1 Basic dosimetric characteristics of the OCTAVIUS 4D MR system

The comprehensive results from the foundational analysis of the OCTAVIUS 4D MR were detailed in a separate manuscript, which is currently undergoing the publication process. This section presents the key findings from the basic tests on detectors, essential for subsequent analyses.

The results demonstrated no significant differences in short-term reproducibility (< 0.2%), dose linearity (< 1%), MU rate dependency (< 0.8%), dose-per-pulse dependency (< 0.4%) and angular dependency (standard deviation < 0.5%). Notably, the OD 1600 MR shows a trend of decreasing sensitivity with increasing MU rate (MU/min), consistent with a previous investigation by PTW. This behavior is attributed to increased recombination with increasing MU rate in the liquid-filled ion chambers.

In both medium and large field sizes, both the OD 1500 MR and OD 1600 MR perform well, with deviations of less than 0.7% from the calculations based on commissioning data in the treatment planning system. For smaller fields, the OD 1500 MR tends to underestimate the relative output factors (< 4.1%) due to volume averaging in its 4.4 mm x 4.4 mm x 3 mm detectors. Conversely, the OD 1600 MR tends to overestimate the relative output factors (< 2.1%) because of the scatter effect in high density liquid.. These results are consistent with results from a previous study on a conventional linac with 0 T [1, 10, 33, 34, 44, 56, 60].

The detector demonstrated the capacity to measure treatment plans with a 5 cm offset. The results of the plan QA of the standard clinical plans are described in the following section.

4.2 Plan QA of standard clinical plans with three '3D' QA phantoms

Figure 1 illustrates the mean GPR for the 3%/3mm criteria across three QA devices. All plans measured with the Delta4+ MR and the OCTAVIUS Detector 1500 MR met the standard threshold of 95%. Increased sensitivity was observed when using the stricter criteria of 2%/2mm with a 90% nominal threshold, maintaining the same trend (figure 2).

For one patient, the historic Delta4+ MR QA measurement was repeated to verify the consistency of the results. The newly measured plan showed a GPR of 98.9%, a DTA of 99.2%, and a mean gamma index of 0.31. These numbers align closely with the earlier results, which recorded a GPR of 98.7%, a DTA of 98.8%, and a mean gamma index of 0.22. The remaining results for the Delta4+ MR were derived from data collected in previous years.

Rectum plans measured with the ArcCHECK-MR demonstrate relatively low results with both types of criterion. The results of gamma comparison analysis with 3%/3mm and the number of beams and segments used to deliver the plan are shown in table 2. In cases where the plans have lower results, an underdose in the high-dose regions was identified (figure 3). The Spearman correlation revealed a strong negative correlation between the results and the number of segments (tables 5, 6).

Patient	Beams	Segments	GPD (3%/3mm)	GPD (2%/2mm)
1	5	48	100	98.8
2	9	69	96.6	94.7
3	5	49	99.9	99.1
4	9	69	91.9	86.1
5	5	50	98.4	93.4

Table 2: The gamma comparison analysis results for rectum plans, using 3%/3mm and 2%/2mm criteria. The plans were measured with the ArcCHECK-MR. The table details the number of beams and segments used in each plan.

Position optimization was applied to the OCTAVIUS Detector 1500 MR, including manual optimization for an individual lymph node plan with a small target region and an offset. The automatic optimization did not improve the result. Without optimization, the GPR was 86.6% for the 2%/2mm criterion. The dose profile (fig. 4) revealed that the detector plate had insufficient resolution for such small field sizes.

4.3 MU errors

The results of the gamma comparison analyses are shown in tables 7, 8. The gamma test was not particularly effective at detecting dose errors in any of the systems in which it was used, especially underdoses in small target regions or highly modulated beams. None of the phantoms were able to detect these errors in plans with small target sizes when using 95% and 90% thresholds. Fig. 5 shows the results for pancreas/liver plans with 3%/3mm criteria, which also demonstrates the typical pattern of the device response. ArcCHECK-MR has greater sensitivity, as shown by decreasing mean GPR and consistent standard deviation. In comparison, Delta4+ MR shows

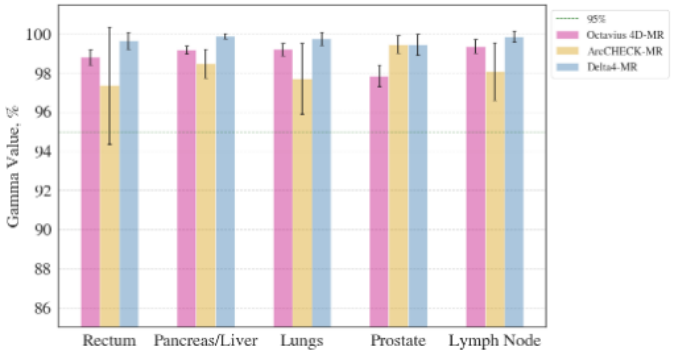

Figure 1: The mean GPR and STD per anatomical site with 3%/3mm criterion and 10% low dose threshold for OCTAVIUS 4D MR, Delta4+ MR, and ArcCHECK-MR. The results are presented as arithmetic averages for each patient group. The nominal value of 95% is shown.

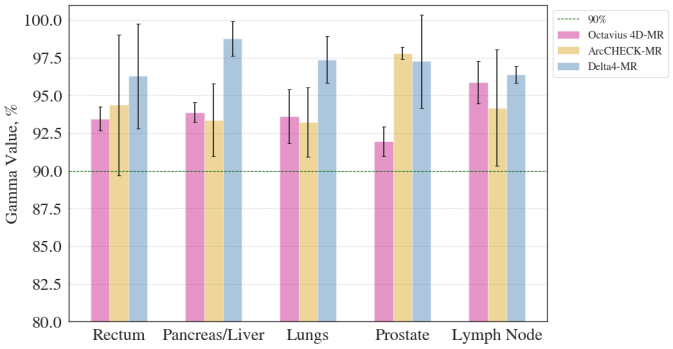

Figure 2: The mean GPR and STD per anatomical site with 2%/2mm criterion and 10% low dose threshold for OCTAVIUS 4D MR, Delta4-MR, and ArcCHECK-MR. The results are presented as arithmetic averages for each patient group. The nominal value of 90% is shown.

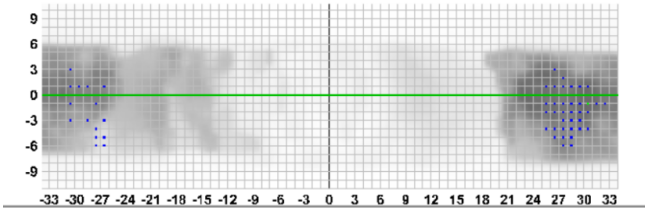

Figure 3: Results of the gamma comparison analysis with 3%/3mm criterion for a rectum plan measured using ArcCHECK-MR. The blue and orange data points delineate the underdose and overdose regions, respectively. The darker the area, the greater the dose it represents.

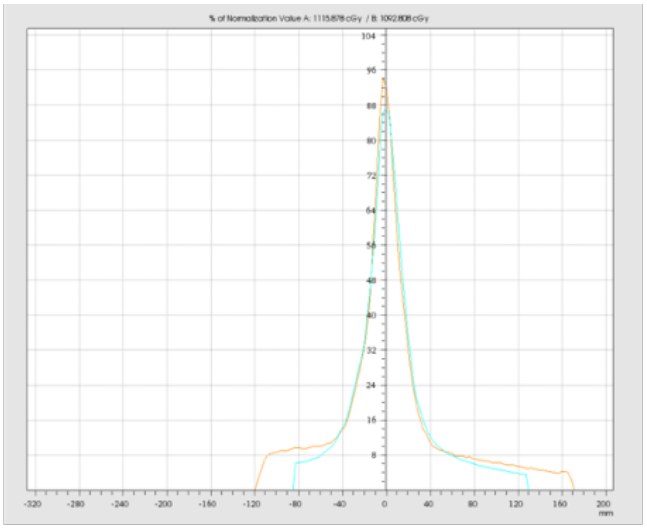

Figure 4: A left-right dose profile for one lymph node plan measured with OCTAVIUS Detector 1500 MR. The orange line shows the planned dose and the blue line shows the measured dose.

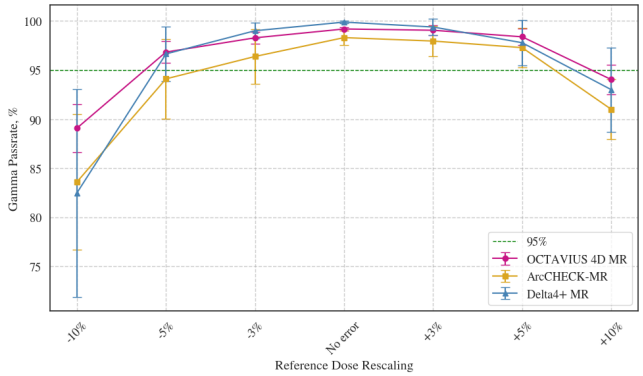

Figure 5: The average GPR for pancreas/liver plans with MU errors across three devices. 3%/3mm criterion was used for the evaluation. The nominal value of 95% is shown. The graph shows an example of a complicated case with a small anatomical region.

less certain results with a higher standard deviation. The GPR calculated in OCTAVIUS 4D MR VeriSoft decreases gradually with a small deviation.

The changes in mean gamma indices between the plans with deliberately introduced errors and standard plans are shown in table 9. Fig. 6 shows the gamma indices changes for pancreas/liver plans with 3%/3mm criterion. Delta4+ MR demonstrates a higher standard deviation across all anatomical sites and higher mean gamma index changes for plans containing MU errors.

The devices show a weaker response in cases involving positive reference data rescaling (underdose) compared to those with negative rescaling (overdose) due to .

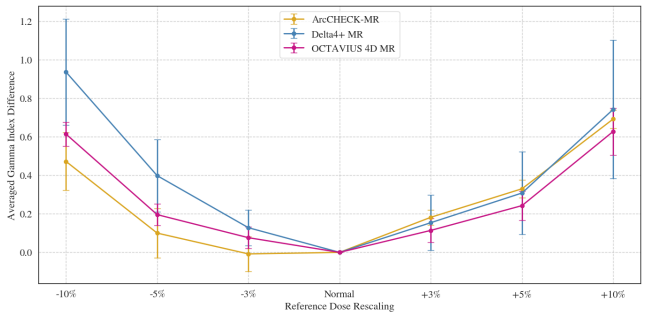

Figure 6: Changes in mean gamma indices between pancreas/liver plans with deliberately introduced MU error and standard plans. The gamma analysis was performed with the 3%/3mm criterion. The results are presented as arithmetic averages for the pancreas/liver group.

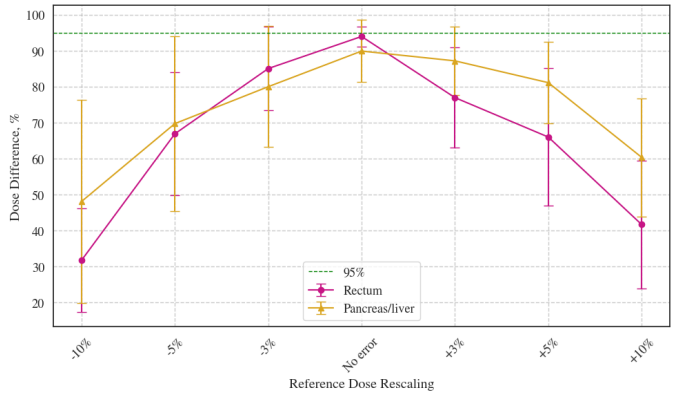

Figure 7: The percentage of points meeting the 3% DD criterion for rectum and pancreas/liver groups for Delta4+ MR. The results are presented as arithmetic averages per anatomical site.

4.3.1 Delta4+ MR dose difference. Within the Delta4+ MR ScandiDos software, DD and median DD were collected. DD shows the percentage of points that are within the DD criterion. This metric demonstrated a significant response and its sensitivity varies according to the target size. With larger targets, the metric responds similarly to overdose and underdose conditions. However, for smaller targets, when the TPS indicates a positive error (underdose), the result is less sensitive than in cases of overdose (fig. 7).

The median DD metric demonstrated a linear correlation with the introduced error magnitude, as shown in fig. 8. A linear regression model was fitted to each set of results by the anatomical site (table 3). For plans without deliberately introduced errors the median dose difference deviated from 'perfect' due to the gradient (penumbra) regions, where the dose agreement is never ideal.

The lymph node plans have a higher standard deviation compared to the other anatomical sites. The example of the gamma comparison analysis with 3%/3mm for a lymph node plan is shown in fig. 9. The yellow area of the dose difference histogram shows

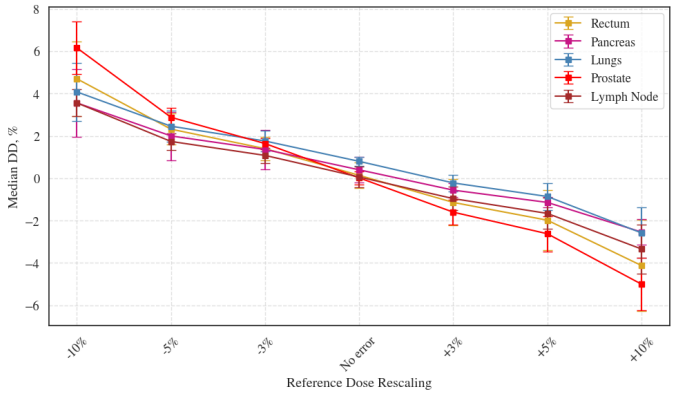

Figure 8: The median dose difference (%) for plans with deliberately introduced MU errors. The results are presented as arithmetic averages per anatomical site. The plans were measured with Delta+ MR.

Anatomical site	R^2	Slope	Intercept
Rectum	0.999	0.44	0.19
Pancreas/liver	0.999	0.31	0.44
Lungs	0.999	0.33	0.78
Prostate	0.995	0.56	0.21
Lymph nodes	0.999	0.34	0.07

Table 3: Median DD approximation as a linear function of the MU error magnitude for Delta4-MR.

multiple spots that exceed the tolerance zone, which is clearly shown on the detector plates. Although the underdose occurs in a high-dose region, the GPD remains at 92.2% (3%/3mm criterion).

4.3.2 ArcCHECK-MR dose difference. Within the ArcCHECK-MR SNC Patient software, mean DD and DD between the points receiving the highest dose were collected. The mean dose difference (%) for plans with deliberately introduced MU errors is shown in fig. 10. The mean DD increases relative to the magnitude of errors. In the case of larger targets such as the rectum and prostate, standard plans have lower mean DD, which grows more rapidly for plans with errors. Lymph node plans have higher initial values but show less dramatic rises.

The percentage difference between the points receiving the highest dose for the ArcCHECK-MR is shown in fig. 11. A linear regression model was fitted to each set of results by the anatomical site (table 4).

4.4 Position errors

The results of the gamma comparison analysis are shown in tables 10, 11. All devices successfully identified positional errors in lung plans. While detection in pancreas/liver plans was slightly less accurate, it was still within acceptable limits. Lymph node errors were also detected by ArcCHECK-MR and Delta4-MR, though the MR-Octavius 4D presented poorer performance. For the prostate and rectum plans, which had a larger size compared to others in this study, positional errors went undetected with Delta4-MR.

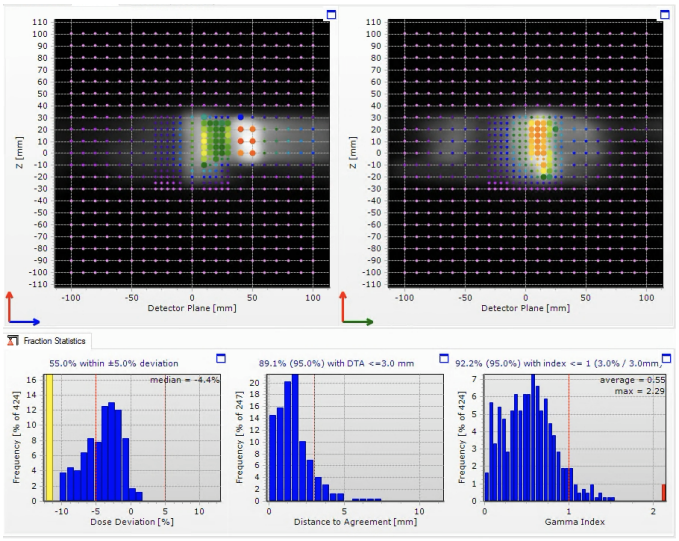

Figure 9: The example of the Lymph node plan measured with the Delta4+ MR where the TPS plan was 'positively' rescaled by 10%. Accompanying this are the DD, DTA, GPD, and median DD metrics with their respective histograms.

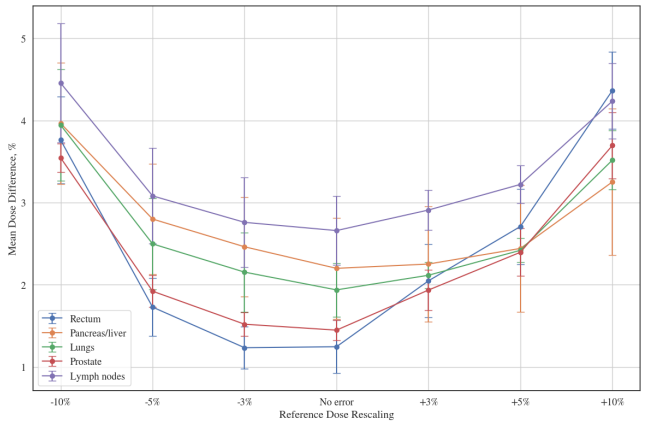

Figure 10: The mean DD (%) for plans with deliberately introduced MU errors. The results are presented as arithmetic averages per anatomical site. The plans were measured with ArcCHECK-MR.

Furthermore, Delta4-MR had the poorest performance for the larger rectum plans, in which all devices struggled to identify shifts (fig. 12).

Overall, ArcCHECK-MR demonstrated the greatest sensitivity to positional shifts under the described conditions. Delta4+ MR had difficulties with larger anatomical sites, whereas Octavius 1500MR exhibited the weakest performance for small target sizes.

The changes in mean gamma indices between the plans with deliberately introduced errors and standard plans are shown in table 12. Fig. 13 shows the gamma indices changes for rectum plans with 3%/3mm criterion. Delta4+ MR and ArcCHECK-MR demonstrate a

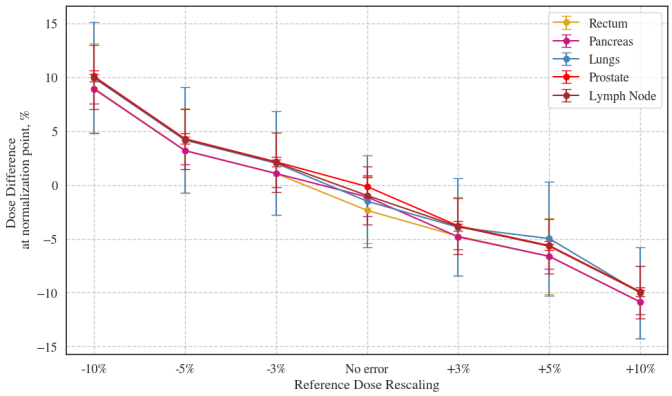

Figure 11: The percentage dose difference between the points receiving the highest dose for plans with deliberately introduced MU errors. The results are presented as arithmetic averages per anatomical site. The plans were measured with ArcCHECK-MR.

Anatomical site	R^2	Slope	Intercept
Rectum	0.994	0.99	-1.6
Pancreas/liver	0.996	0.988	-1.46
Lungs	0.991	0.983	-0.62
Prostate	0.996	0.998	-0.42
Lymph nodes	0.996	0.998	-0.59

Table 4: Linear regression approximating dose difference at the points receiving the highest dose as a function of the MU error magnitude for ArcCHECK-MR.

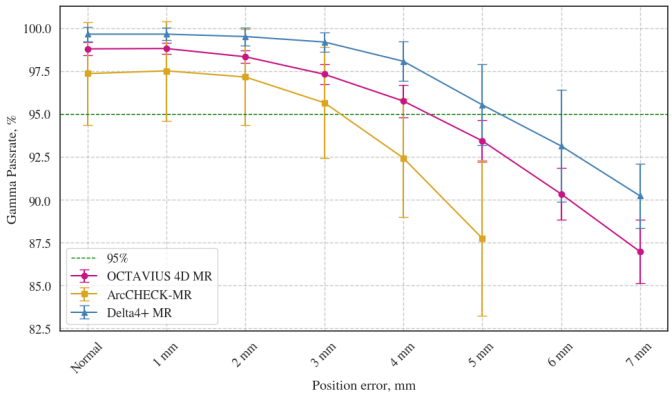

Figure 12: The mean GPR for rectum plans with positional errors across three devices. 3%/3mm criterion was used for the evaluation. The nominal value of 95% is shown. The graph shows an example of a complicated case with a big anatomical region.

high standard deviation across all anatomical sites. ArcCHECK-MR also shows a higher mean gamma index and GPD changes for plans containing position errors.

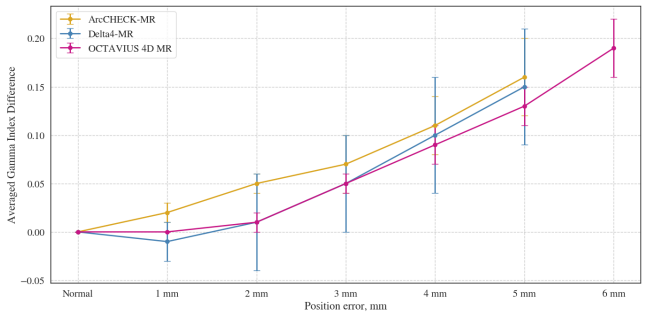

Figure 13: Changes in mean gamma indices between rectum plans with deliberately introduced position error and standard plans. The gamma analysis was performed with the 3%/3mm criterion. The results are presented as arithmetic averages for the rectum group.

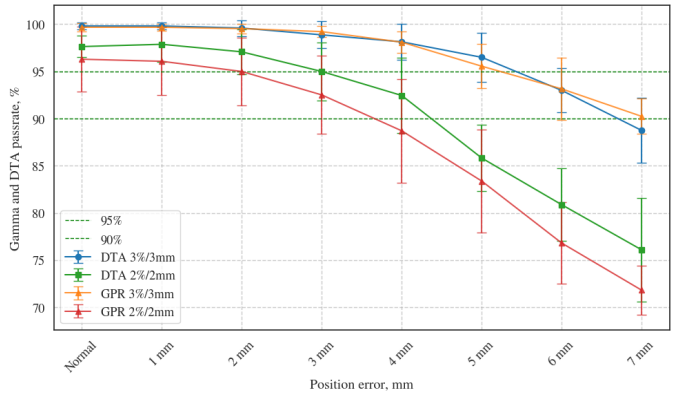

Figure 14: The mean GPR and distance to agreement (DTA) for rectum plans with positional errors for Delta4-MR. 3%/3mm and 2%/2mm criteria were used for the evaluation. The nominal values of 95% and 90% are shown.

4.4.1 Delta4+ MR - distance to agreement. The Delta4+ MR ScandiDos software included DTA metric, which can be valuable for detecting positional errors.

The sensitivity of the DTA and GPD metrics within the Delta4+ MR system was analyzed. Data was collected for rectum plans, which proved to be the most difficult in terms of identifying errors. The same nominal thresholds of 95% and 90% were used. The results are illustrated on the figure 14. The performance of both metrics – DTA and GPD – was similar across both types of criteria, and the DTA outcomes were close to those of the GPD. However, the mean values for GPD, when employing the 2%/2mm criterion, were lower compared to those for DTA.

4.4.2 ArcCHECK-MR - distance to agreement. The ArcCHECK-MR SNC Patient software allows for DTA comparison with two modes 'relative dose' (RD) and 'absolute dose' (AD). In Figure 15, the DTA metrics using both 2%/2mm and 3%/3mm criteria for rectum plans were compared to the GPD metric. The results indicated that the mean value for DTA AD is lower compared to RD. However, the RD

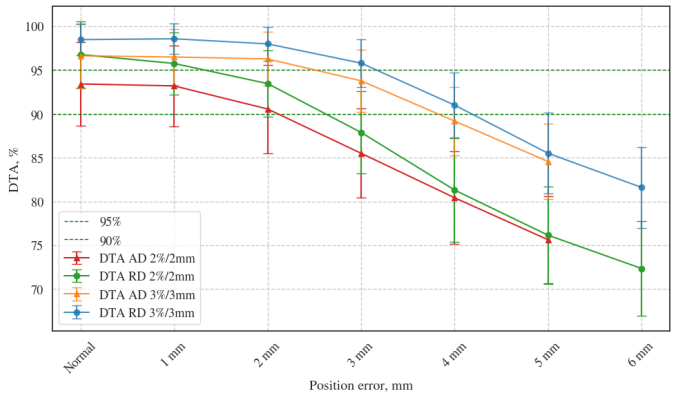

Figure 15: The mean distance to agreement (DTA) in 'relative dose' mode (RD) and 'absolute dose' mode (AD) for rectum plans with positional errors for ArcCHECK-MR. 3%/3mm and 2%/2mm criteria were used for the evaluation. The nominal values of 95% and 90% are shown.

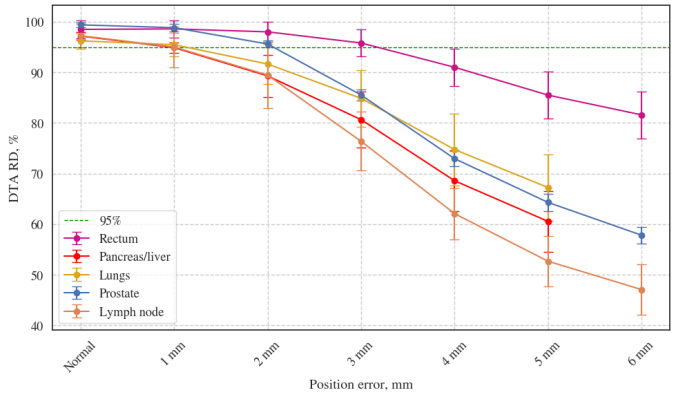

Figure 16: The mean DTA RD for different anatomical sites with positional errors. The results are shown for the ArcCHECK-MR. The 3%/3mm criterion was used for the evaluation.

mode may be more effective for identifying position errors since dose changes had less impact. The results also showed that DTA was more sensitive to position changes than GPD. The DTA RD metric was applied to all other anatomical sites and showed strong capabilities in detecting positional errors when using the 95% and 90% nominal thresholds. The results for the 3%/3mm criterion are shown in fig. 16.

5 DISCUSSION

In this manuscript, the comparison between three QA MR-compatible devices was described. The comparison was based on 25 measured patients' plans and plans containing MU and position error. Before measuring clinical treatment plans, the basic characterization tests of the combination of the OCTAVIUS 4D Phantom MR and the OCTAVIUS Detector 1500 MR chamber array was performed.

The primary difference between the three devices is that the ArcCHECK-MR is designed to measure doses at the beam's entry and exit, with linear interpolation used to assess the center area. In contrast, the Delta4+ MR and OCTAVIUS 4D MR are designed to measure the center region of the phantom, which is often where the target area is located. However, issues may arise with these devices when the target is located in an area not covered by the detector. This is one of the limitations of the Delta4+ MR because the gamma analysis is performed only for the detectors. This means that in some cases the user should rely on the measurements obtained at the detectors region rather than on the target area data. This problem is especially noticeable for MRl settings because of the fixed table position, which requires the user to move the phantom itself in order to match the phantom isocenter with the target region. Volumetric analysis in OCTAVIUS 4D MR allows the estimation of all measured dose inside the phantom which makes the device potentially more suitable for such cases. In such situations, the ArcCHECK-MR might be also more effective since the position of the target area does not affect its use.

A critical challenge in evaluating results is the setup process. UMC Utrecht uses lasers in the MR linac room to place the phantom at the isocenter. While this method of alignment is used, it may not be exact, with errors as small as half a millimeter influencing the results. In this institution, the Delta4+ MR system is used for patient QA, and the typical approach is to apply position optimization to the 'box' plan, and then use the recommended position for clinical plans measured with this setup. A similar method might be used with the OCTAVIUS 4D MR, although in this case, position optimization was done individually for each plan group after the measurements. The ArcCHECK-MR system also has a position optimization tool, however it was not available in our analysis, which indicates any inconsistencies in the results might be due to the misalignment. Relevant observations were found with rectum plans measured using ArcCHECK-MR, where two out of five plans used more beams and segments for delivery and showed lower GPR. An underdose was detected between 160 and 220 degrees, where the beam passes through the table and phantom platform, and especially their edges. The Elekta Unity QA platform used with the ArcCHECK-MR, with an electron density of 1.22, has the potential to significantly contribute to attenuation. Minor misalignments, which are more visible in rectum plans because of the greater number of beams going through the bottom of the phantom, where attenuation effects are more evident, might be a contributing factor. Misalignment was also noticeable in the lymph node plan, which had a small high-dose region with the target offset. When not compensated for the position, both ArcCHECK-MR and OD 1500 MR performed poorly under the more strict criterion. Other factors, however, might also have a role, and a more efficient technique would be to have the same initial alignment of phantoms to avoid these impacts. An important aspect is that alignment should demonstrate the optimal correspondence between the planned and measured doses, which means any inaccuracies may have gone undetected. Typically, this tool is used in clinic for 'box' plan or calibration beam measurements, rather than for patient treatment plans.

The main limitation of the OD 1500 MR is its resolution, which, when combined with the linear interpolation algorithm, might result in overly smoothed 3D dose reconstructions. This problem is particularly visible in the overlays of line profiles [12].

The analysis of plans with MU errors, which simulate dose calibration errors, demonstrated that, under certain conditions, the global gamma percentage difference with a 10%-low dose threshold may not reliably detect dose inaccuracies. Other studies also highlighted the limitations of the analysis in identifying dose errors [38, 40, 54]. In terms of mean gamma indices, all devices demonstrated sensitivity, with gamma indices increasing with the increasing error magnitude. The dose difference, a key component of the gamma index equation, is calculated with respect to the normalization point, which was normally the location receiving the highest dose. When the low-dose zone is large, more low-dose points may pass through the analysis. Furthermore, within low-dose regions, points with significant dose differences but low DTA can provide acceptable gamma values. This could hide dose errors in small targets due to the large low-dose area and additional metrics should be used to identify inaccuracy in the target area and organs at risk. On the other hand, plans with larger targets show more sensitivity to dose changes because of the big high-dose region. The median dose difference and the dose difference at the points of highest dose, collected respectively from the Delta4+ MR and ArcCHECK-MR software, also demonstrated a linear response to the magnitude of errors. Here the impact of the normalization point is more visible demonstrating that a 10% dose adjustment usually leads to mean and median values ranging from 4% to 6%. This happened because the dose was adjusted locally but compared globally. For instance, if the lowest and highest dose points initially received 1 cGy and 50 cGy, and were rescaled by 10%, the new doses were 1.1 cGy and 55 cGy. However, if the normalization is performed to the highest dose, the increase at the low dose point appears minimal, only 0.2%, while at the high dose point, the increase is the full 10%. This discrepancy is shown by the near match of our findings to the scale of our adjustments when the focus is on the dose difference at the highest point. Additional metric for the OCTAVIUS 4D MR that were not considered in this study could be local percentage dose difference. This metric provides information on a slice-by-slice basis, indicating areas of discrepancy.

When investigating plans with introduced position errors, gamma comparison analysis alone was revealed to be insufficient for bigger analomical sites such as rectum. However, all devices responded to the isocenter shifts, and the average mean gamma index increased following the magnitude of errors. Smaller anatomical sites were more sensitive to position changes, and bigger sites were less sensitive due to their big high-dose region. DTA is another metric that is designed to detect positional errors. According to the Delta4 manual, this metric is particularly effective in high-gradient regions, but its effectiveness is reduced in low-dose gradients. When analyzing Delta4+ MR rectum plans, which included bigger anatomical sites, both GPR and DTA produced similar results in terms of the minimum identified error. ArcCHECK DTA in RD mode successfully identified all mistakes that exceeded the tolerance criteria since this metric was less influenced by dose differences.

In reviewing the performance of several devices when considering different error scenarios in this investigation, it was noticeable that the OD 1500 MR had a significantly gradual response than other comparable devices, meaning that the mean GPR decreased

steady with small standard deviations. The ArcCHECK-MR frequently showed a more significant change in the GPR, whereas the Delta4+ MR had a broad range of results and a high standard deviation. These discrepancies could be related to differences in devices' geometries and the space where gamma comparison analysis was performed. The Delta4+ MR showed gamma calculation on two planes where its detectors were located. The algorithm calculated the gamma index for all detectors and limited the search to a distance twice as big as the acceptance criteria. ArcCHECK SNC Patient measured gamma within the detector cylinder surface. The OCTAVIUS 4D considered 3D gamma analysis for every slice and allowed for the volumetric gamma analysis which is statistical overview of the 3D gamma calculation. This method produced more accurate data by including extra points and taking into account all slices inside the phantom. As a consequence, the voxels might 'pass' the volumetric test but 'fail' the gamma analysis on the other devices. This scenario could be reduced to the fact that for ArcCHECK-MR and Delta4+ MR, the 3D gamma calculation was performed for points from 2D planes where the detectors were located, while for the OCTAVIUS 4D MR, points from 3D space were used. While it was expected that this would improve the analysis's efficiency, there were also more 'noisy' voxels involved, which caused the results to change more gradually for deliberate erroneous plans. Besides, another drawback of this algorithm was a significant increase in the amount of data that might result in more complex information organization and computational complexity. As a result, the analytical process took longer to complete.

It should be emphasized as well that the discrepancy between ArcCHECK-MR and other devices might be because ArcCHECK-MR compared wholly different dose map. Delta4+ MR and OCTAVIUS 4D MR measured the dose in the central area, whereas ArcCHECK-MR used diodes placed in a helical pattern to create a separate plane for evaluation. This suggests that the high-dose region may appear differently on the cylinder surface, which was demonstrated in this study to enhance sensitivity in specific circumstances like positioning shift.

The study also contained several limitations by design. First, the simulations might not represent the actual inconsistencies that occur during clinical procedures. The simulated MU errors were designed to mimic potential calibration issues that result in an overdose or underdose with respect to a plan. The simulated positional errors were meant to mimic possible misalignments, however, this study did not account for the fact that any variation in the phantom's position would also have affected the 3D dose reconstruction. Furthermore, the study only investigated a limited number of metrics. One could investigate a wider array of potential inconsistencies in IMRT QA such as realistic MLC positioning errors, missed segments, asymmetric beams etc. A broader range of plans across different anatomical groups could be explored to gain a more comprehensive understanding.

6 CONCLUSION

The new MR compatible OCTAVIUS detectors and the OCTAVIUS 4D Phantom MR are suitable for QA of patient treatment plans in a 1.5T MRI-linac and for measurements with the offset.

All devices demonstrated their capability to produce good results for standard clinical plans. However, minor misalignment during the setup procedure could affect the accuracy of the analysis. The limited resolution and density of the detectors can impact the precision of results in plans with small high-dose regions.

When investigating plans with introduced errors, gamma comparison analysis alone was revealed to be insufficient for MU error detection and should be combined with a dose assessment. Nevertheless, all devices showed a worsening in results for plans with deliberately introduced errors. The responses of the ArcCHECK-MR and Delta4+ MR were more sensitive but also had larger uncertainty.

The geometry of the devices can affect the results. The gamma calculation for ArcCHECK-MR and Delta4+ MR is performed for the diodes location, while OCTAVIUS 4D MR measurements cover a more extensive area of space. Additionally, the measurement surface can cause discrepancies in the results.

7 ACKNOWLEDGEMENT

The authors would like to thank PTW for providing the OCTAVIUS Detector 1500 MR, OCTAVIUS Detector 1600 MR and OCTAVIUS 4D Phantom MR, as well as SunNuclear for providing the ArcCHECK-MR.

REFERENCES

[1] 2017. *Dosimetry of Small Static Fields Used in External Beam Radiotherapy.* Number 483 in Technical Reports Series. INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna. https://www.iaea.org/publications/11075/dosimetry-of-small-static-fields-used-in-external-beam-radiotherapy

[2] Abdulrahman Mohammed Abdulbaqi, Siham Sabah Abdullah, HH Alabed, Nabaa M Alazawy, MJ Al-Musawi, and Ahmed Faris Heydar. 2020. The Correlation of Total MU Number and Percentage Dosimetric Error in Step and Shoot IMRT with Gamma Passing Rate Using OCTAVIUS 4D-1500 Detector Phantom. *Ann Trop Med Public Health* 23, 19 (2020).

[3] Sahaja Acharya, Benjamin W Fischer-Valuck, Rojano Kashani, Parag Parikh, Deshan Yang, Tianyu Zhao, Olga Green, Omar Wooten, H Harold Li, Yanle Hu, et al. 2016. Online magnetic resonance image guided adaptive radiation therapy: first clinical applications. *International Journal of Radiation Oncology* Biology* Physics* 94, 2 (2016), 394–403.

[4] B Allgaier, E Schüle, and J Würfel. 2013. Dose reconstruction in the OCTAVIUS 4D phantom and in the patient without using dose information from the TPS. *PTW White Pap* 913 (2013), 0–7.

[5] Michalis Aristophanous, Yelin Suh, Pai C Chi, Luke J Whittlesey, Scott LaNeave, and Mary K Martel. 2016. Initial clinical experience with ArcCHECK for IMRT/VMAT QA. *Journal of applied clinical medical physics* 17, 5 (2016), 20–33.

[6] S Arumugam, A Xing, G Goozee, and L Holloway. 2013. Detecting VMAT delivery errors: a study on the sensitivity of the ArcCHECK-3D electronic dosimeter. In *Journal of Physics: Conference Series*, Vol. 444. IOP Publishing, 012019.

[7] J. Würfel B. Allgaier, E. Schüle. 2013. Dose reconstruction in the OCTAVIUS 4D phantom and in the patient without using dose information from the TPS. *DocPlayer.net* (2013).

[8] A Bäck. 2015. Quasi 3D dosimetry (EPID, conventional 2D/3D detector matrices). In *Journal of Physics: Conference Series*, Vol. 573. IOP Publishing, 012012.

[9] James L Bedford, Young K Lee, Philip Wai, Christopher P South, and Alan P Warrington. 2009. Evaluation of the Delta4 phantom for IMRT and VMAT verification. *Physics in Medicine & Biology* 54, 9 (2009), N167.

[10] EM Benítez, FJ Casado, S García-Pareja, JA Martín-Viera, C Moreno, and V Parra. 2013. Evaluation of a liquid ionization chamber for relative dosimetry in small and large fields of radiotherapy photon beams. *Radiation measurements* 58 (2013), 79–86.

[11] Vibha Chaswal, Michael Weldon, Nilendu Gupta, Arnab Chakravarti, and Yi Rong. 2014. Commissioning and comprehensive evaluation of the ArcCHECK cylindrical diode array for VMAT pretreatment delivery QA. *Journal of applied clinical medical physics* 15, 4 (2014), 212–225.

[12] PTW THE DOSIMETRY COMPANY. [n.d.]. Code of practice: OCTAVIUS 4D How to start (Monaco). *PTW* ([n. d.]). https://www.ptw-usa.com/en/products/octavius-4d-qa-phantom?downloadfile=1957&type=3451&cHash=293bf15d5176d0984e327f1e58d25326

[13] JHW De Vries, Enrica Seravalli, AC Houweling, Simon J Woodings, Rob van Rooij, Jochem WH Wolthaus, JJW Lagendijk, and Bas W Raaymakers. 2018. Characterization of a prototype MR-compatible Delta4 QA system in a 1.5 tesla MR-linac. *Physics in Medicine & Biology* 63, 2 (2018), 02NT02.

[14] Homan Dehnad, Aart J Nederveen, Uulke A van der Heide, R Jeroen A van Moorselaar, Pieter Hofman, and Jan JW Lagendijk. 2003. Clinical feasibility study for the use of implanted gold seeds in the prostate as reliable positioning markers during megavoltage irradiation. *Radiotherapy and oncology* 67, 3 (2003), 295–302.

[15] Vimal Desai, John Bayouth, Jennifer Smilowitz, and Poonam Yadav. 2021. A clinical validation of the MR-compatible Delta4 QA system in a 0.35 tesla MR linear accelerator. *Journal of Applied Clinical Medical Physics* 22, 4 (2021), 82–91.

[16] Armand Djouguela, Irmgard Grießbach, Dietrich Harder, Ralf Kollhoff, Ndimofor Chofor, Antje Rühmann, Kay Willborn, and Bjoern Poppe. 2008. Dosimetric characteristics of an unshielded p-type Si diode: linearity, photon energy dependence and spatial resolution. *Zeitschrift für Medizinische Physik* 18, 4 (2008), 301–306.

[17] Steven T Ellefson, Wesley S Culberson, Bryan P Bednarz, Larry A DeWerd, and John E Bayouth. 2017. An analysis of the Arc CHECK-MR diode array's performance for ViewRay quality assurance. *Journal of Applied Clinical Medical Physics* 18, 4 (2017), 161–171.

[18] Gary A Ezzell, James M Galvin, Daniel Low, Jatinder R Palta, Isaac Rosen, Michael B Sharpe, Ping Xia, Ying Xiao, Lei Xing, and Cedric X Yu. 2003. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee. *Medical physics* 30, 8 (2003), 2089–2115.

[19] Vladimir Feygelman, Geoffrey Zhang, Craig Stevens, and Benjamin E Nelms. 2011. Evaluation of a new VMAT QA device, or the "X" and "O" array geometries. *Journal of applied clinical medical physics* 12, 2 (2011), 146–168.

[20] AC Houweling, JHW De Vries, J Wolthaus, S Woodings, JGM Kok, B Van Asselen, K Smit, A Bel, JJW Lagendijk, and BW Raaymakers. 2016. Performance of a cylindrical diode array for use in a 1.5 T MR-linac. *Physics in Medicine & Biology* 61, 3 (2016), N80.

[21] Mohammad Hussein, CH Clark, and Andrew Nisbet. 2017. Challenges in calculation of the gamma index in radiotherapy–towards good practice. *Physica Medica* 36 (2017), 1–11.

[22] Mohammad Hussein, Pejman Rowshanfarzad, Martin A Ebert, Andrew Nisbet, and Catharine H Clark. 2013. A comparison of the gamma index analysis in various commercial IMRT/VMAT QA systems. *Radiotherapy and Oncology* 109, 3 (2013), 370–376.

[23] Ina M Jürgenliemk-Schulz, Robbert JHA Tersteeg, Judith M Roesink, Stefan Bijmolt, Christel N Nomden, Marinus A Moerland, and Astrid AC de Leeuw. 2009. MRI-guided treatment-planning optimisation in intracavitary or combined intracavitary/interstitial PDR brachytherapy using tandem ovoid applicators in locally advanced cervical cancer. *Radiotherapy and Oncology* 93, 2 (2009), 322–330.

[24] Jan JW Lagendijk, Bas W Raaymakers, and Marco Van Vulpen. 2014. The magnetic resonance imaging–linac system. In *Seminars in radiation oncology*, Vol. 24. Elsevier, 207–209.

[25] Wolfram U Laub and Tony Wong. 2003. The volume effect of detectors in the dosimetry of small fields used in IMRT. *Medical physics* 30, 3 (2003), 341–347.

[26] Daniel Létourneau, Misbah Gulam, Di Yan, Mark Oldham, and John W Wong. 2004. Evaluation of a 2D diode array for IMRT quality assurance. *Radiotherapy and oncology* 70, 2 (2004), 199–206.

[27] Guangjun Li, Yingjie Zhang, Xiaoqin Jiang, Sen Bai, Guang Peng, Kui Wu, and Qingfeng Jiang. 2013. Evaluation of the ArcCHECK QA system for IMRT and VMAT verification. *Physica medica* 29, 3 (2013), 295–303.

[28] Heng Li, Lei Dong, Lifei Zhang, James N Yang, Michael T Gillin, and X Ronald Zhu. 2011. Toward a better understanding of the gamma index: Investigation of parameters with a surface-based distance method a. *Medical physics* 38, 12 (2011), 6730–6741.

[29] H Harold Li, Vivian L Rodriguez, Olga L Green, Yanle Hu, Rojano Kashani, H Omar Wooten, Deshan Yang, and Sasa Mutic. 2015. Patient-specific quality assurance for the delivery of 60Co intensity modulated radiation therapy subject to a 0.35-T lateral magnetic field. *International Journal of Radiation Oncology* Biology* Physics* 91, 1 (2015), 65–72.

[30] Bin Liang, Bo Liu, Fugen Zhou, Fang-fang Yin, and Qiuwen Wu. 2016. Comparisons of volumetric modulated arc therapy (VMAT) quality assurance (QA) systems: sensitivity analysis to machine errors. *Radiation Oncology* 11, 1 (2016), 1–10.

[31] Daniel A Low and James F Dempsey. 2003. Evaluation of the gamma dose distribution comparison method. *Medical physics* 30, 9 (2003), 2455–2464.

[32] Daniel A Low, William B Harms, Sasa Mutic, and James A Purdy. 1998. A technique for the quantitative evaluation of dose distributions. *Medical physics* 25, 5 (1998), 656–661.

[33] Miljenko Markovic, Sotirios Stathakis, Panayiotis Mavroidis, Ines-Ana Jurkovic, and Nikos Papanikolaou. 2014. Characterization of a two-dimensional liquid-filled ion chamber detector array used for verification of the treatments in radiotherapy. *Medical physics* 41, 5 (2014), 051704.

[34] Chantal Martens, Carlos De Wagter, and Wilfried De Neve. 2001. The value of the LA48 linear ion chamber array for characterization of intensity-modulated beams. *Physics in Medicine & Biology* 46, 4 (2001), 1131.

[35] Conor K McGarry, Barry F O'Connell, Mark WD Grattan, Christina E Agnew, Denise M Irvine, and Alan R Hounsell. 2013. Octavius 4D characterization for flattened and flattening filter free rotational deliveries. *Medical physics* 40, 9 (2013), 091707.

[36] Moyed Miften, Arthur Olch, Dimitris Mihailidis, Jean Moran, Todd Pawlicki, Andrea Molineu, Harold Li, Krishni Wijesooriya, Jie Shi, Ping Xia, et al. 2018. Tolerance limits and methodologies for IMRT measurement-based verification QA: recommendations of AAPM Task Group No. 218. *Medical physics* 45, 4 (2018), e53–e83.

[37] David Mönnich, Jasmin Winter, Marcel Nachbar, Luise Künzel, Simon Boeke, Cihan Gani, Oliver Dohm, Daniel Zips, and Daniela Thorwarth. 2020. Quality assurance of IMRT treatment plans for a 1.5 T MR-linac using a 2D ionization chamber array and a static solid phantom. *Physics in Medicine & Biology* 65, 16 (2020), 16NT01.

[38] Christopher Neilson, Michael Klein, Rob Barnett, and Slav Yartsev. 2013. Delivery quality assurance with ArcCHECK. *Medical Dosimetry* 38, 1 (2013), 77–80.

[39] Benjamin E Nelms, Maria F Chan, Geneviève Jarry, Matthieu Lemire, John Lowden, Carnell Hampton, and Vladimir Feygelman. 2013. Evaluating IMRT and VMAT dose accuracy: practical examples of failure to detect systematic errors when applying a commonly used metric and action levels. *Medical physics* 40, 11 (2013), 111722.

[40] Benjamin E Nelms, Heming Zhen, and Wolfgang A Tomé. 2011. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Medical physics* 38, 2 (2011), 1037–1044.

[41] American Association of Physicists in Medicine et al. 2009. TG-119 IMRT Commissioning Tests Instructions for Planning, Measurement and Analysis. *Med. Phys* 36 (2009), 5359–5373.

[42] Bipasha Pal, Angshuman Pal, Santanu Bag, Md Abbas Ali, Suresh Das, Soura Palit, Papai Sarkar, Suman Mallik, Jyotirup Goswami, Sayan Das, et al. 2021. Comparative performance analysis of 2D and 3D gamma metrics for patient specific QA in VMAT using Octavius 4D with 2D-Array 1500. *Physica Medica* 91 (2021), 18–27.

[43] E Pappas, TG Maris, A Papadakis, F Zacharopoulou, J Damilakis, N Papanikolaou, and N Gourtsoyiannis. 2006. Experimental determination of the effect of detector size on profile measurements in narrow photon beams. *Medical physics* 33, 10 (2006), 3700–3710.

[44] Bjoern Poppe, Arne Blechschmidt, Armand Djouguela, Ralf Kollhoff, Antje Rubach, Kay C Willborn, and Dietrich Harder. 2006. Two-dimensional ionization chamber arrays for IMRT plan verification. *Medical physics* 33, 4 (2006), 1005–1015.

[45] Richard Pötter, Petra Georg, Johannes CA Dimopoulos, Magdalena Grimm, Daniel Berger, Nicole Nesvacil, Dietmar Georg, Maximilian P Schmid, Alexander Reinthaller, Alina Sturdza, et al. 2011. Clinical outcome of protocol based image (MRI) guided adaptive brachytherapy combined with 3D conformal radiotherapy with or without chemotherapy in patients with locally advanced cervical cancer. *Radiotherapy and Oncology* 100, 1 (2011), 116–123.

[46] Alexander JE Raaijmakers, Bas W Raaymakers, and Jan JW Lagendijk. 2005. Integrating a MRI scanner with a 6 MV radiotherapy accelerator: dose increase at tissue–air interfaces in a lateral magnetic field due to returning electrons. *Physics in Medicine & Biology* 50, 7 (2005), 1363.

[47] Bas W Raaymakers, IM Jürgenliemk-Schulz, GH Bol, M Glitzner, ANTJ Kotte, B Van Asselen, JCJ De Boer, JJ Bluemink, SL Hackett, MA Moerland, et al. 2017. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Physics in Medicine & Biology* 62, 23 (2017), L41.

[48] Dhanabalan Rajasekaran, Prakash Jeevanandam, Prabakar Sukumar, Arulpandiyan Ranganathan, Samdevakumar Johnjothi, and Vivekanandan Nagarajan. 2014. A study on correlation between 2D and 3D gamma evaluation metrics in patient-specific quality assurance for VMAT. *Medical Dosimetry* 39, 4 (2014), 300–308.

[49] Ram Sadagopan, J BenComo, R Martin, P Balter, S Vedam, and G Nilsson. 2007. Characterisation, commissioning and evaluation of Delta4 IMRT QA system. *Med Phys* 34, 6 (2007), 2560.

[50] Ramaswamy Sadagopan, Jose A Bencomo, Rafael L Martin, Gorgen Nilsson, Thomas Matzen, and Peter A Balter. 2009. Characterization and clinical evaluation of a novel IMRT quality assurance system. *Journal of Applied Clinical Medical Physics* 10, 2 (2009), 104–119.

[51] Masahide Saito, Naoki Sano, Yuki Shibata, Kengo Kuriyama, Takafumi Komiyama, Kan Marino, Shinichi Aoki, Kazunari Ashizawa, Kazuya Yoshizawa, and Hiroshi Onishi. 2018. Comparison of MLC error sensitivity of various commercial devices for VMAT pre-treatment quality assurance. *Journal of Applied Clinical Medical Physics* 19, 3 (2018), 87–93.

[52] Jin Ho Song, Hun-Joo Shin, Chul Seung Kay, and Seok Hyun Son. 2015. Dosimetric verification by using the ArcCHECK system and 3DVH software for various target sizes. *PLoS One* 10, 3 (2015), e0119937.

[53] Cassandra Stambaugh, Daniel Opp, Stuart Wassrman, Geoffrey Zhang, and Vladimir Feygelman. 2014. Evaluation of semiempirical VMAT dose reconstruction on a patient dataset based on biplanar diode array measurements. *Journal of Applied Clinical Medical Physics* 15, 2 (2014), 169–180.
[54] Jennifer M Steers and Benedick A Fraass. 2016. IMRT QA: selecting gamma criteria based on error detection sensitivity. *Medical physics* 43, 4 (2016), 1982–1994.
[55] Jennifer M Steers and Benedick A Fraass. 2021. IMRT QA and gamma comparisons: The impact of detector geometry, spatial sampling, and delivery technique on gamma comparison sensitivity. *Medical Physics* 48, 9 (2021), 5367–5381.
[56] Tenzin Sonam Stelljes, A Harmeyer, Julia Reuter, Hui Khee Looe, Ndimofor Chofor, Dietrich Harder, and Bjoern Poppe. 2015. Dosimetric characteristics of the novel 2D ionization chamber array OCTAVIUS Detector 1500. *Medical physics* 42, 4 (2015), 1528–1537.
[57] Robert Timmerman, Rebecca Paulus, James Galvin, Jeffrey Michalski, William Straube, Jeffrey Bradley, Achilles Fakiris, Andrea Bezjak, Gregory Videtic, David Johnstone, et al. 2010. Stereotactic body radiation therapy for inoperable early stage lung cancer. *Jama* 303, 11 (2010), 1070–1076.
[58] Yuuki Tomiyama, Fujio Araki, Takeshi Oono, and Kazunari Hioki. 2014. Three-dimensional gamma analysis of dose distributions in individual structures for IMRT dose verification. *Radiological physics and technology* 7 (2014), 303–309.
[59] Patrizia Urso, Rita Lorusso, Luca Marzoli, Daniela Corletto, Paolo Imperiale, Annalisa Pepe, and Lorenzo Bianchi. 2018. Practical application of Octavius®-4D: characteristics and criticalities for IMRT and VMAT verification. *Journal of Applied Clinical Medical Physics* 19, 5 (2018), 517–524.
[60] Ann Van Esch, Katarzyna Basta, Marie Evrard, Michel Ghislain, Francois Sergent, and Dominique P Huyskens. 2014. The Octavius1500 2D ion chamber array and its associated phantoms: Dosimetric characterization of a new prototype. *Medical physics* 41, 9 (2014), 091708.
[61] Simon J Woodings, JJ Bluemink, JHW De Vries, Yury Niatsetski, Bob van Veelen, Joost Schillings, Jan GM Kok, Jochem WH Wolthaus, Sara L Hackett, Bram van Asselen, et al. 2018. Beam characterisation of the 1.5 T MRI-linac. *Physics in Medicine & Biology* 63, 8 (2018), 085015.
[62] Aitang Xing, Sankar Arumugam, Shrikant Deshpande, Armia George, Philip Vial, Lois Holloway, and Gary Goozee. 2015. Evaluation of 3D Gamma index calculation implemented in two commercial dosimetry systems. In *Journal of Physics: Conference Series*, Vol. 573. IOP Publishing, 012054.
[63] Tai Keung Yeung, Karen Bortolotto, Scott Cosby, Margaret Hoar, and Ernst Lederer. 2005. Quality assurance in radiotherapy: evaluation of errors and incidents recorded over a 10 year period. *Radiotherapy and oncology* 74, 3 (2005), 283–291.
[64] Heming Zhen, Benjamin E Nelms, and Wolfgang A Tomé. 2011. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. *Medical physics* 38, 10 (2011), 5477–5489.

A APPENDIX

Anatomical site	ArcCHECK-MR 3%/3mm				Octavius 1500MR 3%/3mm				Delta4+ MR 3%/3mm			
	Beams		Segments		Beams		Segments		Beams		Segments	
	Corr.	P-Value	Corr.	P-Value	Corr.	P-Value	Corr.	P-Value	Corr.	P-Value	Corr.	P-Value
Rectum	−0.866	0.058	−0.975	0.005	0.000	1.000	0.205	0.741	0.889	0.044	0.684	0.203
Pancreas	−0.949	0.051	−0.800	0.200	0.316	0.684	0.600	0.400	−0.707	0.293	−0.894	0.106
Lungs	0.447	0.450	−0.410	0.493	0.057	0.927	0.526	0.362	0.884	0.047	0.216	0.727
Prostate	nan	nan	0.132	0.833	nan	nan	−0.821	0.089	nan	nan	−0.947	0.014
Lymph Nodes	−0.577	0.308	−0.616	0.269	0.289	0.638	−0.051	0.935	−0.612	0.272	−0.725	0.165
All plans	−0.489	0.015	−0.152	0.478	0.474	0.019	−0.300	0.154	0.182	0.381	−0.417	0.043

Table 5: The results from the Spearman correlation analysis for the ArcCHECK-MR, Delta4+ MR, and Octavius 1500MR. P-values below 0.05 show a statistically significant correlation between the GPD with 3%/3mm criterion and the amount of beams and segments.

Anatomical site	ArcCHECK-MR 2%/2mm				Octavius 1500MR 2%/2mm				Delta4+ MR 2%/2mm			
	Beams		Segments		Beams		Segments		Beams		Segments	
	Corr.	P-Value	Corr.	P-Value	Corr.	P-Value	Corr.	P-Value	Corr.	P-Value	Corr.	P-Value
Rectum	−0.577	0.308	−0.718	0.172	−0.866	0.058	−0.975	0.005	0.866	0.058	0.564	0.322
Pancreas	−0.632	0.368	−0.400	0.600	−0.316	0.684	0.000	1.000	−0.316	0.684	−0.400	0.600
Lungs	0.783	0.118	0.103	0.870	−0.224	0.718	0.564	0.322	0.630	0.254	0.526	0.362
Prostate	nan	nan	0.205	0.741	nan	nan	−0.921	0.026	nan	nan	−0.948	0.014
Lymph Nodes	−0.289	0.638	−0.462	0.434	0.289	0.638	−0.051	0.935	−0.296	0.628	−0.500	0.391
All plans	−0.528	0.008	0.021	0.923	0.257	0.225	−0.404	0.050	0.009	0.968	−0.236	0.266

Table 6: The results from the Spearman correlation analysis for the ArcCHECK-MR, Delta4+ MR, and Octavius 1500MR. P-values below 0.05 show a statistically significant correlation between the GPD with 2%/2mm criterion and the amount of beams and segments.

	+10%	+5%	+3%	Normal	-3%	-5%	-10%
Rectum							
ArcCHECK-MR	63.2 ± 3.6	83.9 ± 5.0	91.5 ± 4.7	97.4 ± 3.0	98.9 ± 1.4	96.6 ± 2.6	73.5 ± 6.4
Delta4-MR	66.4 ± 16.6	82.3 ± 15.8	89.7 ± 10.9	99.6 ± 0.4	97.4 ± 2.3	86.1 ± 7.0	51.8 ± 15.2
Octavius 1500MR	68.8 ± 6.9	86.4 ± 5.0	94.7 ± 2.9	98.8 ± 0.4	94.2 ± 1.9	86.1 ± 1.2	68.5 ± 2.5
Pancreas/liver							
ArcCHECK-MR	91.0 ± 3.1	97.3 ± 2.0	98.0 ± 1.5	98.3 ± 0.8	96.4 ± 2.8	94.1 ± 4.0	83.6 ± 6.9
Delta4-MR	93.0 ± 4.3	97.8 ± 2.3	99.4 ± 0.8	99.9 ± 0.2	99.0 ± 0.8	96.6 ± 2.7	82.5 ± 10.6
Octavius 1500MR	94.0 ± 1.5	98.4 ± 0.8	99.1 ± 0.5	99.2 ± 0.2	98.3 ± 0.6	96.8 ± 1.1	89.1 ± 2.5
Lungs							
ArcCHECK-MR	85.5 ± 5.4	94.1 ± 2.5	96.2 ± 1.8	97.7 ± 1.8	97.3 ± 1.2	94.4 ± 2.6	81.5 ± 3.9
Delta4-MR	90.8 ± 10.2	98.6 ± 2.2	99.7 ± 0.4	99.7 ± 0.3	97.1 ± 2.3	92.1 ± 5.5	72.8 ± 16.4
Octavius 1500MR	92.8 ± 2.9	98.7 ± 0.2	99.4 ± 0.2	99.2 ± 0.3	97.2 ± 1.4	94.5 ± 2.4	82.4 ± 6.1
Prostate							
ArcCHECK-MR	79.0 ± 1.5	93.0 ± 2.5	97.2 ± 1.2	99.4 ± 0.4	99.4 ± 0.6	98.2 ± 0.8	85.3 ± 5.6
Delta4-MR	73.0 ± 11.9	78.1 ± 9.8	89.8 ± 5.9	99.4 ± 0.5	95.5 ± 1.8	81.5 ± 7.6	60.7 ± 13.1
Octavius 1500MR	89.4 ± 1.5	97.1 ± 0.6	98.8 ± 0.2	97.8 ± 0.5	92.1 ± 2.0	85.7 ± 3.3	68.6 ± 4.4
Lymph node							
ArcCHECK-MR	87.9 ± 7.1	94.6 ± 3.5	96.4 ± 2.5	98.1 ± 1.5	98.0 ± 1.9	97.0 ± 1.8	87.7 ± 4.1
Delta4-MR	93.6 ± 5.7	99.0 ± 0.9	99.5 ± 0.6	99.9 ± 0.3	99.9 ± 0.2	99.1 ± 1.3	88.3 ± 7.0
Octavius 1500MR	95.0 ± 2.5	98.5 ± 0.9	99.0 ± 0.8	98.8 ± 1.0	97.6 ± 1.3	95.7 ± 2.0	86.2 ± 6.3

Table 7: Gamma percentage difference with standard deviation for clinical plans with MU errors. The analysis was performed with the 3%/3mm criterion. The results are presented as arithmetic averages per anatomical site.

	+10%	+5%	+3%	Normal	-3%	-5%	-10%
Rectum							
ArcCHECK-MR	52.0 ± 3.9	70.9 ± 5.3	82.4 ± 6.4	94.4 ± 4.7	96.3 ± 2.4	89.3 ± 5.5	59.9 ± 6.4
Delta4-MR	49.2 ± 16.2	72.7 ± 16.6	82.0 ± 15.0	96.3 ± 3.5	89.3 ± 6.8	61.3 ± 17.7	30.3 ± 9.6
Octavius 1500MR	47.1 ± 7.1	71.4 ± 6.9	82.6 ± 6.0	92.6 ± 1.6	83.5 ± 2.4	73.7 ± 1.9	53.4 ± 2.6
Pancreas/liver							
ArcCHECK-MR	79.5 ± 6.1	89.7 ± 5.1	93.0 ± 3.7	93.4 ± 2.4	88.7 ± 4.7	83.7 ± 3.8	71.5 ± 7.9
Delta4-MR	77.8 ± 6.6	94.3 ± 3.3	97.0 ± 2.0	98.8 ± 1.2	94.9 ± 3.3	86.7 ± 8.2	60.4 ± 16.5
Octavius 1500MR	76.6 ± 3.5	89.5 ± 2.6	92.6 ± 1.9	93.8 ± 0.7	91.4 ± 1.4	87.6 ± 2.2	72.3 ± 3.9
Lungs							
ArcCHECK-MR	71.8 ± 8.2	86.8 ± 3.7	90.5 ± 2.9	93.6 ± 2.4	90.6 ± 2.9	86.0 ± 4.0	67.9 ± 4.4
Delta4-MR	77.0 ± 16.0	94.5 ± 4.6	97.6 ± 2.0	97.4 ± 1.6	89.0 ± 6.1	78.9 ± 11.4	49.5 ± 15.8
Octavius 1500MR	73.8 ± 6.1	90.1 ± 0.7	93.5 ± 0.9	93.6 ± 1.6	88.4 ± 2.6	82.4 ± 3.9	62.9 ± 7.0
Prostate							
ArcCHECK-MR	66.8 ± 2.9	81.8 ± 3.5	90.2 ± 3.8	97.9 ± 0.5	97.5 ± 1.6	93.9 ± 3.5	71.2 ± 6.4
Delta4-MR	56.9 ± 14.4	71.4 ± 12.1	78.7 ± 9.6	97.3 ± 3.1	84.7 ± 6.3	64.4 ± 17.2	43.6 ± 13.4
Octavius 1500MR	71.4 ± 2.7	88.0 ± 1.9	92.1 ± 1.2	90.6 ± 1.3	79.7 ± 3.6	71.4 ± 4.0	54.1 ± 3.0
Lymph node							
ArcCHECK-MR	74.4 ± 8.3	85.7 ± 6.2	88.6 ± 5.4	94.2 ± 3.9	93.7 ± 5.0	90.2 ± 5.8	74.8 ± 5.9
Delta4-MR	77.5 ± 14.0	95.0 ± 3.8	98.2 ± 1.6	99.4 ± 0.6	98.2 ± 1.7	93.7 ± 2.4	69.3 ± 11.6
Octavius 1500MR	78.7 ± 5.7	90.3 ± 4.3	92.5 ± 3.9	92.9 ± 3.5	89.2 ± 4.0	84.6 ± 5.5	68.9 ± 8.2

Table 8: Gamma percentage difference with standard deviation for clinical plans with MU errors. The analysis was performed with the 2%/2mm criterion. The results are presented as arithmetic averages per anatomical site.

	+10%	+5%	+3%	-3%	-5%	-10%
Rectum						
ArcCHECK-MR	0.69 ± 0.05	0.33 ± 0.05	0.18 ± 0.04	−0.01 ± 0.09	0.10 ± 0.13	0.47 ± 0.15
Delta4+ MR	0.74 ± 0.36	0.31 ± 0.22	0.15 ± 0.14	0.13 ± 0.09	0.40 ± 0.19	0.94 ± 0.27
Octavius 1500MR	0.63 ± 0.12	0.24 ± 0.08	0.11 ± 0.06	0.08 ± 0.06	0.20 ± 0.06	0.61 ± 0.06
Pancreas/Liver						
ArcCHECK-MR	0.15 ± 0.06	0.04 ± 0.05	0.01 ± 0.03	0.03 ± 0.06	0.08 ± 0.09	0.25 ± 0.17
Delta4+ MR	0.27 ± 0.03	0.08 ± 0.03	0.02 ± 0.03	0.08 ± 0.06	0.17 ± 0.09	0.48 ± 0.17
Octavius 1500MR	0.19 ± 0.02	0.06 ± 0.02	0.02 ± 0.01	0.02 ± 0.02	0.06 ± 0.03	0.21 ± 0.06
Lungs						
ArcCHECK-MR	0.27 ± 0.12	0.08 ± 0.07	0.03 ± 0.04	0.04 ± 0.03	0.09 ± 0.04	0.29 ± 0.07
Delta4+ MR	0.23 ± 0.20	0.02 ± 0.09	−0.02 ± 0.05	0.14 ± 0.07	0.26 ± 0.12	0.59 ± 0.24
Octavius 1500MR	0.23 ± 0.07	0.06 ± 0.03	0.01 ± 0.01	0.08 ± 0.03	0.15 ± 0.03	0.38 ± 0.06
Prostate						
ArcCHECK-MR	0.43 ± 0.03	0.20 ± 0.02	0.11 ± 0.02	0.01 ± 0.06	0.09 ± 0.08	0.35 ± 0.08
Delta4+ MR	0.77 ± 0.24	0.36 ± 0.13	0.19 ± 0.08	0.16 ± 0.07	0.39 ± 0.10	1.06 ± 0.26
Octavius 1500MR	0.24 ± 0.02	0.03 ± 0.02	−0.01 ± 0.01	0.10 ± 0.01	0.20 ± 0.02	0.51 ± 0.03
Lymph node						
ArcCHECK-MR	0.23 ± 0.11	0.10 ± 0.05	0.05 ± 0.02	0.00 ± 0.02	0.05 ± 0.05	0.21 ± 0.09
Delta4+ MR	0.30 ± 0.13	0.11 ± 0.07	0.05 ± 0.04	0.06 ± 0.04	0.14 ± 0.06	0.41 ± 0.11
Octavius 1500MR	0.16 ± 0.03	0.04 ± 0.01	0.01 ± 0.01	0.02 ± 0.03	0.06 ± 0.05	0.22 ± 0.10

Table 9: The magnitude of changes in the gamma index between plans with deliberately introduced MU error and standard plans. The gamma analysis was performed with the 3%/3mm criterion. The results are presented as arithmetic averages per anatomical site.

	Normal	1 mm	2 mm	3 mm	4 mm	5 mm	6 mm	7 mm	10 mm
Rectum									
ArcCHECK-MR	97.4 ± 3.0	97.5 ± 2.9	97.2 ± 2.8	95.7 ± 3.3	92.4 ± 3.4	87.8 ± 4.5	87.4 ± 2.5	nan	nan
Delta4+ MR	99.7 ± 0.4	99.7 ± 0.4	99.5 ± 0.5	99.2 ± 0.6	98.1 ± 1.1	95.5 ± 2.4	93.1 ± 3.3	90.2 ± 1.9	nan
Octavius 1500MR	98.8 ± 0.4	98.8 ± 0.3	98.3 ± 0.4	97.3 ± 0.6	95.8 ± 1.0	93.4 ± 1.2	90.3 ± 1.5	87.0 ± 1.9	76.3 ± 2.9
Pancreas/Liver									
ArcCHECK-MR	98.3 ± 0.7	97.2 ± 1.3	91.9 ± 3.5	84.4 ± 5.9	73.7 ± 8.7	63.1 ± 9.6	nan	nan	nan
Delta4+ MR	99.9 ± 0.2	99.8 ± 0.2	98.2 ± 2.2	93.2 ± 7.2	83.0 ± 12.1	75.1 ± 13.6	65.4 ± 13.8	nan	nan
Octavius 1500MR	99.2 ± 0.2	99.2 ± 0.2	98.6 ± 0.5	96.9 ± 1.2	93.4 ± 2.1	88.1 ± 2.9	81.8 ± 2.9	75.9 ± 2.5	63.2 ± 1.4
Lungs									
ArcCHECK-MR	97.7 ± 1.8	97.3 ± 2.1	94.2 ± 3.0	88.7 ± 4.4	81.2 ± 7.2	70.9 ± 9.6	nan	nan	nan
Delta4+ MR	99.7 ± 0.3	99.5 ± 0.8	98.7 ± 0.7	92.4 ± 2.2	80.8 ± 5.5	69.8 ± 9.0	64.1 ± 10.3	nan	nan
Octavius 1500MR	99.2 ± 0.3	99.0 ± 0.3	98.1 ± 0.6	95.9 ± 1.0	91.2 ± 1.4	84.9 ± 1.7	78.7 ± 1.8	74.0 ± 2.6	60.1 ± 3.0
Prostate									
ArcCHECK-MR	99.5 ± 0.5	99.1 ± 0.3	97.3 ± 1.6	88.9 ± 1.5	76.7 ± 2.8	68.1 ± 2.9	nan	nan	nan
Delta4+ MR	99.4 ± 0.5	99.5 ± 0.5	98.8 ± 2.0	96.6 ± 4.2	92.2 ± 5.1	86.4 ± 4.7	81.0 ± 4.8	nan	nan
Octavius 1500MR	97.8 ± 0.6	97.8 ± 0.5	97.3 ± 0.6	95.7 ± 1.2	92.6 ± 1.9	87.8 ± 2.5	82.4 ± 2.8	77.4 ± 3.0	65.6 ± 3.5
Lymph node									
ArcCHECK-MR	98.1 ± 1.5	96.2 ± 3.1	91.5 ± 4.0	79.7 ± 4.9	64.2 ± 5.8	52.2 ± 7.1	nan	nan	nan
Delta4+ MR	99.9 ± 0.3	99.9 ± 0.3	99.7 ± 0.5	96.9 ± 2.1	85.6 ± 7.1	75.8 ± 11.4	67.0 ± 9.1	nan	nan
Octavius 1500MR	98.8 ± 1.0	99.1 ± 0.3	98.9 ± 0.4	97.5 ± 1.6	92.8 ± 3.6	87.7 ± 6.2	80.3 ± 6.0	71.1 ± 4.2	57.1 ± 4.7

Table 10: Gamma percentage difference with standard deviation for clinical plans with position errors. The analysis was performed with the 3%/3mm criterion. The results are presented as arithmetic averages per anatomical site.

	Normal	1 mm	2 mm	3 mm	4 mm	5 mm	6 mm	7 mm	10 mm
Rectum									
ArcCHECK-MR	94.4 ± 4.7	94.3 ± 4.5	92.1 ± 5.2	87.7 ± 5.0	82.5 ± 5.5	77.6 ± 5.9	78.0 ± 3.8	nan	nan
Delta4+ MR	96.3 ± 3.5	96.1 ± 3.6	95.0 ± 3.6	92.5 ± 4.1	88.7 ± 5.5	83.4 ± 5.4	76.8 ± 4.3	71.8 ± 2.6	nan
Octavius 1500MR	92.6 ± 1.6	92.6 ± 1.6	91.2 ± 1.8	88.9 ± 2.1	85.3 ± 2.4	80.8 ± 2.7	76.0 ± 3.2	71.5 ± 3.7	59.9 ± 4.6
Pancreas/Liver									
ArcCHECK-MR	93.4 ± 2.4	88.0 ± 4.5	79.5 ± 6.8	65.9 ± 9.5	56.2 ± 10.9	46.9 ± 8.7	nan	nan	nan
Delta4+ MR	98.8 ± 1.2	96.4 ± 3.4	86.9 ± 10.7	74.8 ± 13.4	63.4 ± 13.4	52.2 ± 12.2	46.2 ± 11.0	nan	nan
Octavius 1500MR	93.8 ± 0.7	93.7 ± 1.2	90.8 ± 2.4	85.0 ± 3.7	77.1 ± 4.1	69.1 ± 3.4	62.4 ± 2.7	57.1 ± 2.0	46.2 ± 0.8
Lungs									
ArcCHECK-MR	93.6 ± 2.4	91.3 ± 3.5	83.2 ± 5.1	73.6 ± 8.0	62.8 ± 9.7	55.3 ± 10.2	nan	nan	nan
Delta4+ MR	97.4 ± 1.6	95.7 ± 1.6	87.3 ± 3.2	72.3 ± 7.5	62.0 ± 10.4	54.2 ± 11.7	48.9 ± 11.8	nan	nan
Octavius 1500MR	93.6 ± 1.6	92.6 ± 1.4	88.7 ± 1.4	81.5 ± 1.8	73.2 ± 2.1	65.7 ± 2.0	59.3 ± 1.8	55.1 ± 2.7	43.2 ± 3.2
Prostate									
ArcCHECK-MR	97.8 ± 0.4	94.7 ± 1.5	84.7 ± 1.8	71.0 ± 2.0	59.4 ± 3.0	52.7 ± 2.9	nan	nan	
Delta4+ MR	97.3 ± 3.1	95.7 ± 5.3	92.2 ± 7.3	85.4 ± 7.3	77.6 ± 7.0	70.2 ± 6.3	63.0 ± 5.6	nan	nan
Octavius 1500MR	90.6 ± 1.3	90.1 ± 1.1	87.6 ± 1.5	82.6 ± 2.0	76.0 ± 2.4	69.5 ± 2.8	63.9 ± 3.0	59.3 ± 3.1	49.5 ± 3.1
Lymph node									
ArcCHECK-MR	94.2 ± 3.9	87.5 ± 5.2	73.6 ± 5.7	55.1 ± 4.5	42.3 ± 4.7	34.9 ± 6.0	nan	nan	nan
Delta4+ MR	99.4 ± 0.6	99.0 ± 0.8	93.4 ± 2.8	78.1 ± 8.9	64.4 ± 10.4	56.7 ± 9.8	50.5 ± 8.4	nan	nan
Octavius 1500MR	92.9 ± 3.5	93.5 ± 1.5	92.5 ± 1.8	86.9 ± 5.0	75.8 ± 6.7	69.4 ± 5.8	61.3 ± 5.6	53.0 ± 4.0	40.6 ± 4.2

Table 11: Gamma percentage difference with standard deviation for clinical plans with position errors. The analysis was performed with the 2%/2mm criterion. The results are presented as arithmetic averages per anatomical site.

	1 mm	2 mm	3 mm	4 mm	5 mm	6 mm
Rectum						
ArcCHECK-MR	0.02 ± 0.01	0.05 ± 0.01	0.07 ± 0.03	0.11 ± 0.03	0.16 ± 0.04	nan
Delta4-MR	−0.01 ± 0.02	0.01 ± 0.05	0.05 ± 0.05	0.10 ± 0.06	0.15 ± 0.06	nan
Octavius 1500MR	0.000	0.01 ± 0.01	0.05 ± 0.01	0.09 ± 0.02	0.13 ± 0.02	0.19 ± 0.03
Pancreas/liver						
ArcCHECK-MR	0.04 ± 0.01	0.13 ± 0.02	0.23 ± 0.03	0.35 ± 0.04	0.48 ± 0.06	nan
Delta4-MR	0.05 ± 0.03	0.15 ± 0.05	0.26 ± 0.08	0.38 ± 0.10	0.50 ± 0.12	0.61 ± 0.15
Octavius 1500MR	0.00 ± 0.02	0.03 ± 0.02	0.09 ± 0.02	0.15 ± 0.02	0.23 ± 0.03	0.31 ± 0.03
Lungs						
ArcCHECK-MR	0.03 ± 0.03	0.11 ± 0.05	0.18 ± 0.07	0.28 ± 0.09	0.38 ± 0.11	nan
Delta4-MR	0.05 ± 0.02	0.14 ± 0.03	0.25 ± 0.05	0.37 ± 0.07	0.49 ± 0.10	0.60 ± 0.13
Octavius 1500MR	0.02 ± 0.02	0.06 ± 0.03	0.12 ± 0.03	0.19 ± 0.03	0.26 ± 0.03	0.35 ± 0.04
Prostate						
ArcCHECK-MR	0.06 ± 0.01	0.15 ± 0.01	0.25 ± 0.01	0.37 ± 0.03	0.48 ± 0.03	nan
Delta4-MR	0.04 ± 0.01	0.10 ± 0.02	0.17 ± 0.03	0.25 ± 0.04	0.33 ± 0.04	0.41 ± 0.05
Octavius 1500MR	0.000	0.04 ± 0.02	0.08 ± 0.02	0.13 ± 0.03	0.19 ± 0.04	0.25 ± 0.05
Lymph node						
ArcCHECK-MR	0.09 ± 0.04	0.19 ± 0.04	0.32 ± 0.04	0.47 ± 0.05	0.62 ± 0.10	nan
Delta4-MR	0.06 ± 0.02	0.15 ± 0.04	0.27 ± 0.07	0.39 ± 0.09	0.51 ± 0.11	0.83 ± 0.54
Octavius 1500MR	−0.01 ± 0.02	0.03 ± 0.04	0.08 ± 0.06	0.15 ± 0.08	0.24 ± 0.09	0.34 ± 0.11

Table 12: The magnitude of changes in the gamma index between plans with deliberately introduced position error and standard plans. The gamma analysis was performed with the 3%/3mm criterion. The results are presented as arithmetic averages per anatomical site.