Patient QA systems for rotational radiation therapy: A comparative experimental study with intentional errors

Anna Fredh, Jonas Bengtsson Scherman, and Lotte S. Fog
Department of Radiation Oncology, Radiation Medicine Research Center, Blegdamsvej 9, DK-2100, Copenhagen, Denmark

Per Munck af Rosenschöld[a)]
Department of Radiation Oncology, Radiation Medicine Research Center, Blegdamsvej 9, DK-2100, Copenhagen, Denmark and The Niels Bohr Institute, University of Copenhagen, DK-2100, Copenhagen Denmark

(Received 26 September 2012; revised 27 November 2012; accepted for publication 27 December 2012; published 28 February 2013)

Purpose: The purpose of the present study was to investigate the ability of commercial patient quality assurance (QA) systems to detect linear accelerator-related errors.
Methods: Four measuring systems (Delta$^{4®}$, OCTAVIUS®, COMPASS, and Epiqa^{TM}) designed for patient specific quality assurance for rotational radiation therapy were compared by measuring four clinical rotational intensity modulated radiation therapy plans as well as plans with introduced intentional errors. The intentional errors included increasing the number of monitor units, widening of the MLC banks, and rotation of the collimator. The measurements were analyzed using the inherent gamma evaluation with 2% and 2 mm criteria and 3% and 3 mm criteria. When applicable, the plans with intentional errors were compared with the original plans both by 3D gamma evaluation and by inspecting the dose volume histograms produced by the systems.
Results: There was considerable variation in the type of errors that the various systems detected; the failure rate for the plans with errors varied between 0% and 72%. When using 2% and 2 mm criteria and 95% as a pass rate the Delta$^{4®}$ detected 15 of 20 errors, OCTAVIUS® detected 8 of 20 errors, COMPASS detected 8 of 20 errors, and Epiqa^{TM} detected 20 of 20 errors. It was also found that the calibration and measuring procedure could benefit from improvements for some of the patient QA systems.
Conclusions: The various systems can detect various errors and the sensitivity to the introduced errors depends on the plan. There was poor correlation between the gamma evaluation pass rates of the QA procedures and the deviations observed in the dose volume histograms. © *2013 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4788645]

I. INTRODUCTION

It is common practice to perform a measurement and/or calculation test in order to verify that the treatment intended for the radiation therapy patient is correct.[1] The patient specific quality assurance (QA) not only varies between different clinics but also depends on the type of treatment given. As new treatment techniques are implemented new methods of quality assurance need to be invented and investigated. The relatively new treatment technique inversely optimized rotational therapy[2] delivers dose to the patient with varying gantry rotation speed, dose rate, and MLC position during treatment. This complex treatment delivery demands sophisticated QA not only for the commissioning of the new treatment technique[3] but QA is also typically performed for each patient plan. The patient QA procedure offers an end-to-end test that is commonly thought to or expected to be catching clinically relevant errors occurring at any point in the whole treatment chain. Therefore, patient dosimetry QA should ideally find errors related to, e.g., CT-geometry, treatment planning computational dosimetry errors (intermittent or systematic), data transmission errors, and mechanical errors at the treatment machine, such as the MLC calibration,[1] and is therefore considered a crucial part of the QA procedure at most radiotherapy clinics. Many different ways of performing QA for rotational therapy have been suggested, for example: using Monte Carlo calculations for verification of the treatment plans,[4,5] log files of the linac,[6,7] electronic portal imager device (EPID) of the linac,[8,9] gel dosimetry,[10,11] and other combinations of ion chamber measurements, film measurements, and commercial systems.[12–14] Many of the commercial systems have previously been described and/or evaluated for intensity modulated radiation therapy (IMRT) treatments and sometimes also for rotational therapy. Chandraraj *et al.*[15] and Masi *et al.*[16] have compared commercial systems for rotational therapy. Given the amount of time that radiation therapy clinics world-wide invest in pretreatment QA of IMRT and rotational IMRT, it is of interest to evaluate if QA systems in current use actually can detect dosimetry and geometry errors that are the foundation for the test performed. The purpose of the present study was to critically compare four commercial systems designed to perform patient specific QA for IMRT and rotational therapy and examine the ability of visualizing intentionally introduced errors. We believe that this study will help clinical physicist and researchers to be aware of the type of errors that can be seen using the QA systems

studied. The knowledge of the characteristics of the particular patient QA system may assist clinical physicist in the assessment of the strengths and weaknesses of a complete QA program.

II. MATERIALS AND METHODS

II.A. Description of the treatment plans

Four rotational IMRT (RapidArc®) plans were created in treatment planning system (TPS) (Eclipse version 8.9, Varian Medical Systems, Inc, Palo Alto, USA), using the analytical anisotropic algorithm (AAA) and 0.25 cm grid size for calculation. The treatment sites in this study were prostate, head and neck (H&N), and brain, all were planned with 6 MV photons and daily IGRT were used. The prostate case had a planning target volume (PTV) of 122 cm^3 and prescribed dose of 2 Gy × 39. The plan had jaw sizes of 10.3 × 10.3 cm^2, 470 MU, and one arc of 300°. The H&N case was an oropharynx cancer with prescribed dose of 2 Gy × 34 to primary tumor plus margin (volume = 381 cm^3). The high risk elective volume (294 cm^3) was treated to 60 Gy and the low risk elective volume (391 cm^3) to 50 Gy. The plan had jaw sizes of 19.9 × 20.2 cm^2, 308 MU, and one arc of 359.8°. The brain cases were high-grade glioma with prescribed dose of 2 Gy × 30, both with a PTV volume of 318 cm^3. The first plan had jaw sizes of 10.4 × 10.4 cm^2, 301 MU, and a 250° arc. The second plan had jaw sizes of 9.8 × 9.8 cm^2, 259 MU, and a 359.8° arc. Instead of making many plans of the same type, different types of plans were chosen since the complexity of the plans will be different. The head and neck plan had three different dose levels and many organs at risk (OAR) that had to be accounted for when creating the plan, while the prostate and brain cases had quite spherical targets but with different sizes, they also had fewer OAR.

From copies of the original plans, five new plans were created which each contained an introduced error, all based on realistic machine calibration errors. The following modifications were introduced; a widening of the MLC bank with 2 and 4 mm divided equally on each side, 3% increased number of monitor units (referred to as "3% dose error" from now on), and a collimator rotation error of 2° and 5°, respectively. The errors were above or on the limit of tolerance levels of regular QA checks as defined by AAPM TG 40.[17]

II.B. Measuring systems and techniques

The patient plans were delivered using a Clinac iX (Varian Medical Systems, Inc) accelerator equipped with a Millenium 120 MLC (0.5 cm leaf width 20 cm centrally and 1 cm leaf width in the outer 20 cm of the field). All plans, both the original plans and the plans with intentional errors, were delivered onto the systems described separately below and the measurement was recorded and saved in the respective software. Patient QA system calibration was performed using the software version available to us and in accordance with the manual provided by the manufacturer updated at the time of measurement (summer 2010). Originally, Sun Nuclear lent us a system (ArcCHECK) to be included in the study,[18] but because of data corruption in that version, it had to be removed from the study.

II.B.1. Delta^{4}®

(ScandiDos AB, Uppsala) More details on the Delta^{4}® system can be found in publications by Korreman *et al.*[12] and Bedford *et al.*[19] This is a system consisting of two orthogonal detector arrays placed in a cylindrical PMMA phantom and an inclinometer to measure the gantry angle. The detector arrays are 20 × 20 cm^2 with a total of 1069 diodes spaced 10 mm apart except for the 6 × 6 cm^2 area in the middle where the spacing is 5 mm. The detector volume is 0.00004 cm^3. As recommended by the manufacturer each of the original plans were recalculated on a virtually constructed CT-scan based on drawings of the phantom, provided by the manufacturer, and with the couch included in the calculation. Prior to measurements four 10 × 10 cm^2 fields were delivered and used for dose calibration and for a more precise alignment of the phantom. Measurements and analysis was done with software version 1.00.0064, where the gamma evaluation was calculated on the two planes.

II.B.2. OCTAVIUS®

(PTW Freiburg) More details on the seven29 2D-array and the OCTAVIUS® system can be found in Spezi *et al.*[20] and Van Esch *et al.*[21] This system consists of the seven29 2D-array with 729 ion chambers, which is inserted into an octagonal polystyrene phantom. The size of the ion chambers is 0.125 cm^3 and the center-to-center distance is 10 mm. The original plans were recalculated on a CT-scanning of the phantom performed in the clinic and with the couch included in the calculation. Dose calibration prior to measurements was done by delivering a 10 × 10 cm^2 field and providing the system the delivered dose. Measurements and analysis was done using the Verisoft version 4.1.0.18 software, where the 3D dose matrix from the TPS was compared with the measurement in one plane, which by the manufacturer is called 3D gamma evaluation.

II.B.3. COMPASS

(IBA dosimetry GmbH) More details on the COMPASS system can be found in Boggula *et al.*[22] This system consists of a gantry angle sensor and a MatriXX^{Evolution} ion chamber array with 1020 ion chambers with a volume of 0.08 cm^3 and spaced 7.62 mm apart center-to-center. Both are mounted on the gantry and measure the gantry angle and the output perpendicular to the radiation during the rotation, respectively. From the TPS a DICOM-export of the patient CT-data, the structures, the plan, and the dose matrix was performed. The software (COMPASS version 2.0.7.0) uses the measured data and the imported plan file to calculate the dose in the imported patient CT-data using a collapsed cone superposition algorithm. The gamma evaluation is calculated in the 3D patient volume. The detector was mounted with source-detector-distance of 100 cm in a gantry holder, rotating with the

collimator. Therefore, the system is not able to detect collimator rotation error, and this is considered to be part of another quality assurance procedure. Prior to measurements, detector commissioning, including dose calibration and detector alignment, was done.

II.B.4. Epiqa™

(EPIdos s.r.o.) This system uses the EPID of the accelerator, in this case an aS1000 (Varian Medical Systems, Inc) with a resolution of 0.78 mm, to acquire an integrated image of the total field. In the TPS a verification plan was calculated on a cubic water phantom with the beam orthogonally incident to one of the sides. The plan and the dose plane from 1.5 cm depth were exported and the 2D gamma evaluation was done in the Epiqa™ software version 1.3.3. More details on the Epiqa™ system can be found in Nicolini *et al.*[23] and Fogliata *et al.*[24]

II.C. Analysis

The impact of the errors in the modified plans was analyzed by comparing the original plan (the plan without errors) with the corresponding plans with intentional error in the TPS. Comparison of dose distributions was made by a gamma evaluation in three dimensions using an in-house made MATLAB (MathWorks Inc.) program with the same criteria as in the measurements. Dose volume histograms (DVH) were compared for the gross tumor volume (GTV), PTV, and one of the organs at risk (OAR). A common tolerance level for the gamma index evaluation is a pass rate of 95%, and an error was considered detected when the gamma failure rate was higher than 5%.

The gamma evaluation tool[25] was available for analysis of the results in all systems and was used with dose difference of 3% and distance to agreement of 3 mm (denoted as "3%/3 mm", from now on) as well as 2%/2 mm criteria. The global gamma index based on absolute data was used, the maximum measured dose was used as the reference dose for normalization, and a cutoff dose at 10% was used (including only measurement points with dose higher than 10% of the maximum measured dose). For the COMPASS a 10% isodose structure was created to correspond to the 10% cutoff, in the other systems. For Epiqa™ the setting of a cutoff level was not possible. Instead the "CIAO feature", which includes the total of the field size bounded by the outermost MLC limits, was used in the evaluation.

In order to see if the different systems were more or less likely to identify problems for the 20 plans with introduced errors, we created tables with binary outcomes of gamma index test for the systems with the criteria of 95% pass rate for 2% and 2 mm. The total detection rate was calculated as the ratio of the number of tests with a fail rate larger than 5% and the total number of tests. McNemar's test was used comparing the Delta$^{4®}$ system-–which was considered reference, as it was our current practice—with OCTAVIUS®, COMPASS, and Epiqa™. A p-value of 0.05 or less was considered statistically significant. The eight plans with collimator errors were both included and excluded in the analysis for the COMPASS system, which by design do not detect collimator rotation errors.

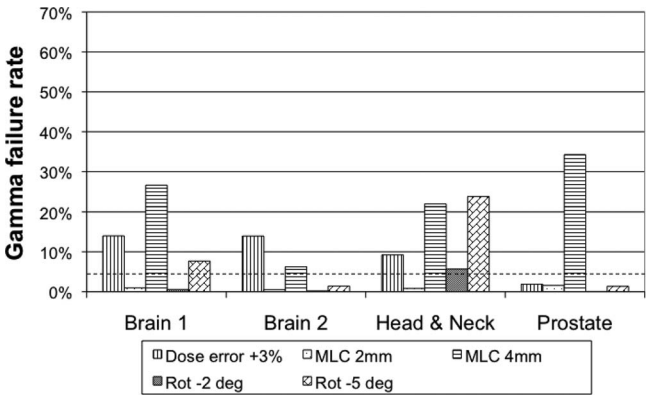

FIG. 1. Results of the 3D gamma index comparison of the plans with 2%/2 mm gamma criteria using the in-house made MATLAB program for plans calculated in the treatment planning system. The errors that are above the threshold (the dashed line) can be expected to be found using the patient QA systems, and these errors were dissimilar for the various plans.

III. RESULTS

For an easier comparison and better overview, the results presented are percentage of the measuring points with a gamma value above 1, here called the gamma failure rate.

In the dose distribution comparison with 3%/3 mm criteria all the plans had a failure rate of less than 5%, except the H&N plan with 5° collimator rotation, which had a failure rate of 12%. The collimator rotation error for the H&N plan was, therefore, the only error expected to be detected by the measuring systems using the 3%/3 mm criteria, under the assumption that (1) the plans are adequately sampled using the patient QA systems and that (2) the TPS calculations are correct. The result of the gamma evaluation of the plan comparison using 2%/2 mm criteria is presented in Fig. 1, where it can be seen that the 2 mm widening of the MLC and the collimator rotation of 2° were the errors with least impact on the plans, they had a failure rate of less than 5% in all cases except for the 2° collimator rotation error of the H&N plan. It was also clear that the impact of an error was plan-dependent. For the 5° collimator rotational error we saw some of the smallest failure rate compared to the original plan for the second brain case and the prostate case, whereas for the H&N plan that was the error with the largest failure rate.

In Table I, the minimum, maximum, and mean percentage dose difference between the plans with error and the original plans calculated using the TPS for GTV, PTV, and one OAR are presented. For the brain, H&N, and prostate case the OAR presented are the chiasm, the spinal cord, and the rectum, respectively. The results in the dose distribution comparison differ somewhat from the gamma evaluation comparison, for example, in the first brain case there was a larger difference compared to the original plan for the 2 mm MLC error than for the 5° collimator rotational error, whereas in the gamma evaluation comparison it was the opposite.

TABLE I. Percentage difference of the minimum dose, the maximum dose, and the mean dose between the plans with errors and the original plans for three structures. In the table, the plans with a gamma failure rate above 5% (2%/2 mm criteria; cf. Fig. 1) are written in italic font. The OAR refers to the chiasm for the brain cases, the spinal cord for the head & neck case, and the rectum for the prostate case.

		GTV (%)			PTV (%)			OAR (%)		
		Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Brain	*Dose +3%*	*3.1*	*3.4*	*3.2*	*2.4*	*3.5*	*3.2*	*0.6*	*2.8*	*1.4*
Case 1	MLC 2 mm	1.1	0.9	1.1	2.5	1.3	1.3	2.4	4.4	3.3
	MLC 4 mm	*1.9*	*1.8*	*2.3*	*4.1*	*2.5*	*2.5*	*4.7*	*7.1*	*6.6*
	Rot −2°	−0.1	−0.2	−0.2	−0.7	−0.5	−0.2	0.1	0.9	0.6
	Rot −5°	*−0.2*	*−0.2*	*−0.1*	*−1.2*	*2.4*	*−0.2*	*0.7*	*3.3*	*1.9*
Brain	*Dose +3%*	*3.0*	*3.1*	*3.1*	*2.5*	*3.2*	*3.0*	*0.7*	*2.8*	*1.3*
Case 2	MLC 2 mm	0.9	1.1	0.7	1.3	0.8	0.8	1.0	1.0	1.3
	MLC 4 mm	*1.4*	*2.3*	*1.4*	*2.2*	*1.4*	*1.5*	*1.7*	*1.7*	*2.2*
	Rot −2°	−0.4	−0.1	−0.1	0.8	0.5	0.0	0.1	1.0	0.1
	Rot −5°	−0.9	0.1	−0.2	1.3	0.9	−0.1	−0.1	2.6	0.5
Head & Neck	*Dose +3%*	*2.6*	*3.3*	*3.2*	*5.1*	*2.8*	*4.3*	*0.1*	*2.1*	*1.5*
	MLC 2 mm	2.3	1.3	1.8	1.9	1.0	1.6	0.2	1.9	1.1
	MLC 4 mm	*4.1*	*3.4*	*3.5*	*3.0*	*3.1*	*3.1*	*0.4*	*4.0*	*2.3*
	Rot −2°	*−1.5*	*1.5*	*0.1*	*−0.9*	*2.3*	*0.5*	*0.1*	*0.3*	*−0.2*
	Rot −5°	*−4.1*	*6.5*	*0.2*	*−8.5*	*9.5*	*1.0*	*0.2*	*7.7*	*0.0*
Prostate	Dose +3%	3.0	3.3	3.1	2.8	3.3	3.1	0.1	3.2	1.4
	MLC 2 mm	1.5	1.3	1.7	2.3	1.4	1.8	0.2	1.5	2.4
	MLC 4 mm	*2.7*	*3.2*	*3.4*	*4.5*	*3.6*	*3.5*	*0.4*	*2.9*	*4.9*
	Rot −2°	−0.2	1.0	0.2	−0.4	1.2	0.1	0.0	0.3	0.0
	Rot −5°	−2.2	3.8	0.4	−2.5	4.7	0.1	0.0	0.8	0.0

Additionally, in this comparison some of the errors in the prostate case have the same impact on the dose to structures as in the other plans, which cannot be seen in the 3D gamma comparison.

The impact of the errors in the prostate plan appears to be more prominent when looking at the DVH than for the gamma evaluation, as can be seen in Fig. 2. This was also a typical example of the DVH appearance when comparing the original plan with the plans with errors. The solid lines represent the original plan and the dashed lines represent the plans with errors: (a) 3% dose error, (b) 2 mm widening of MLC, (c) 4 mm widening of MLC, and (d) 5° collimator rotation. The DVH for the 2° collimator rotation has been omitted since there were even smaller differences than on the 5° collimator rotation DVH. One can see that even though the gamma comparison of the plans gave a small failure rate, there was an obvious difference when looking at the DVH, which may indicate that a detection of such an error would be preferable even though the gamma evaluation comparison by using the in house made MATLAB program of the plans does not detect the error.

In Figs. 3–6 the results of the QA measurements using the different systems are presented. On the Y-axis is the percent gamma failure rate, each pattern of the bars represents the original plan or one of the plans with introduced errors. For the results in Figs. 3–5 and 6(a) the gamma criteria was 2%/2 mm and in Figs. 3–5 and 6(b) the gamma criteria was 3%/3 mm.

The results of the first brain case are shown in Fig. 3. The 3% dose error or the 4 mm MLC error had the highest failure rate for all systems except for the OCTAVIUS® system where the 5° collimator rotation error had the highest failure rate. The second braincase (Fig. 4) was the case where gamma failure rate generally was the lowest. When using 3%/3 mm criteria each of the systems could find only one or none of the errors. For 2%/2 mm criteria more errors were detected but not as many as for the other cases. The H&N case had generally the highest gamma failure rate for the plans with the errors, and the systems also detected more of the errors, see Fig. 5. The Epiqa™ system detected all of the errors for this case for both of the criteria, but it had a high failure rate for the original plan with the 2%/2 mm criteria. For the prostate case (Fig. 6) the 4 mm MLC error was the error with the highest failure rate for all the systems.

The detection rates of the systems, using 2%/2 mm criteria, were 75%, 40%, 40%, and 100% for the Delta$^{4®}$, OCTAVIUS®, COMPASS, and Epiqa™ systems, respectively. Excluding the two angular tests for the COMPASS system elevates the detection rate to 67%. Using McNemar's test, which is a nonparametric test for a difference in proportion in two paired dichotomous samples, the detection rate of the Delta$^{4®}$ system was higher than the OCTAVIUS® ($p = 0.039$) and the COMPASS systems ($p = 0.016$), and somewhat lower than the Epiqa™ system ($p = 0.063$). Removing the plans with collimator error from the analysis the

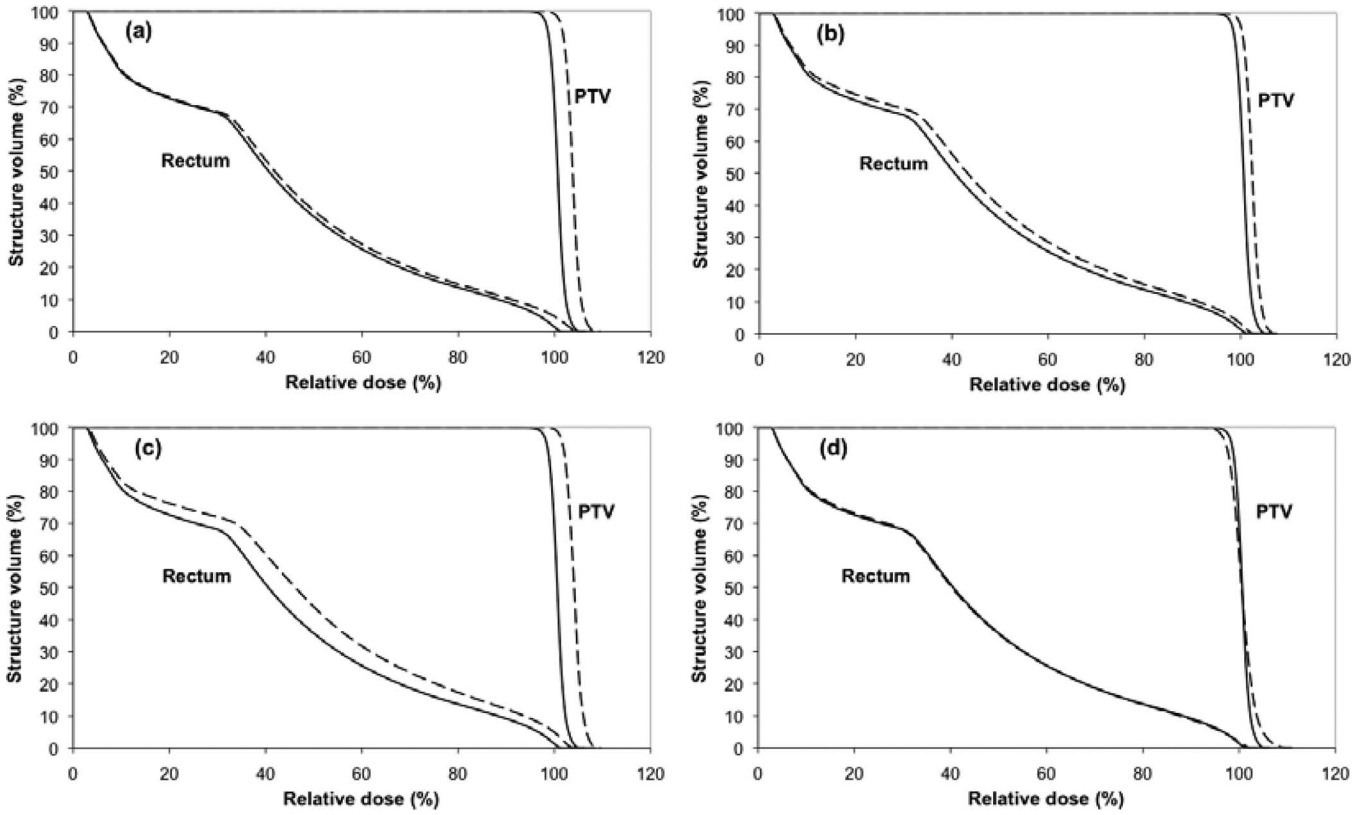

FIG. 2. DVHs from the Eclipse TPS for the PTV and the rectum of the prostate case. The solid lines represent the original plan and the dashed lines represent the following errors: (a) 3% dose error, (b) 2 mm widening of MLC, (c) 4 mm widening of MLC, and (d) 5° collimator rotation.

detection rates of the Delta$^{4®}$ and COMPASS systems were comparable ($p = 0.25$).

IV. DISCUSSION

In the present study, a set of four systems for patient QA has been evaluated. Measurements were performed on a set of four rotational IMRT plans with and without introduced intentional errors. The errors were selected based on machine errors that were deemed to be probable to occur at a radiotherapy clinic at some point. Other errors might occur as well and therefore the present study shall not be seen as a complete and final evaluation of patient QA systems. Of the errors selected not all of them introduced significant dosimetric deviations in the treatment plans made for real patients in our clinic. Only 10 of the total 20 were found when comparing dose distributions of the plans with the introduced error with the original ones using the 3D gamma evaluation with 2%/2 mm criteria. But for some of the introduced errors there was an obvious difference when comparing DVH of the original plans and the plans with errors, even though the error was not found using the gamma evaluation. Poor correlation between gamma evaluation and comparison of DVH has also been discussed by Nelms *et al.*[26] and Zhen *et al.*[27]

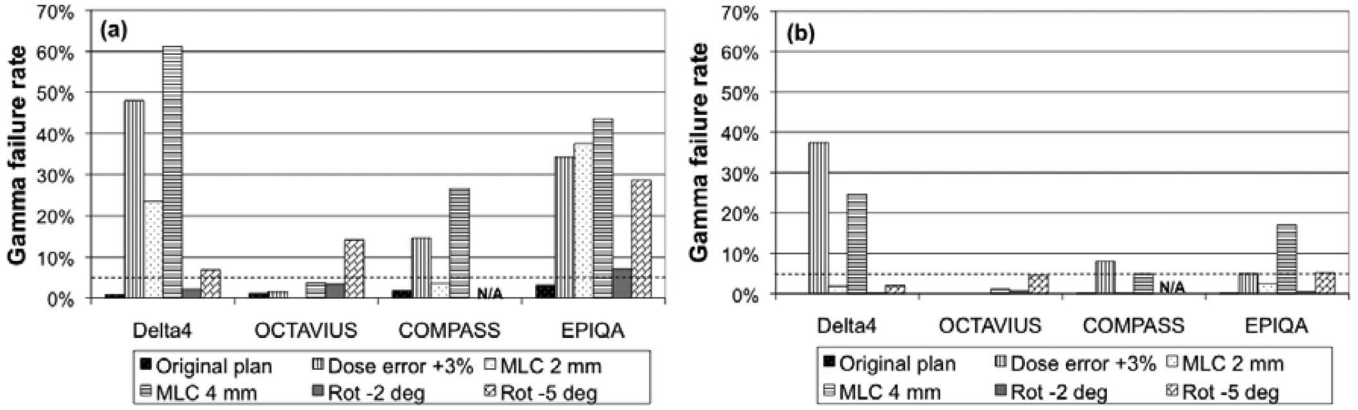

FIG. 3. The results for the first brain case, in (a) the gamma criteria of 2%/2 mm was used and in (b) 3%/3 mm was used. The dashed line represents a failure rate of 5%.

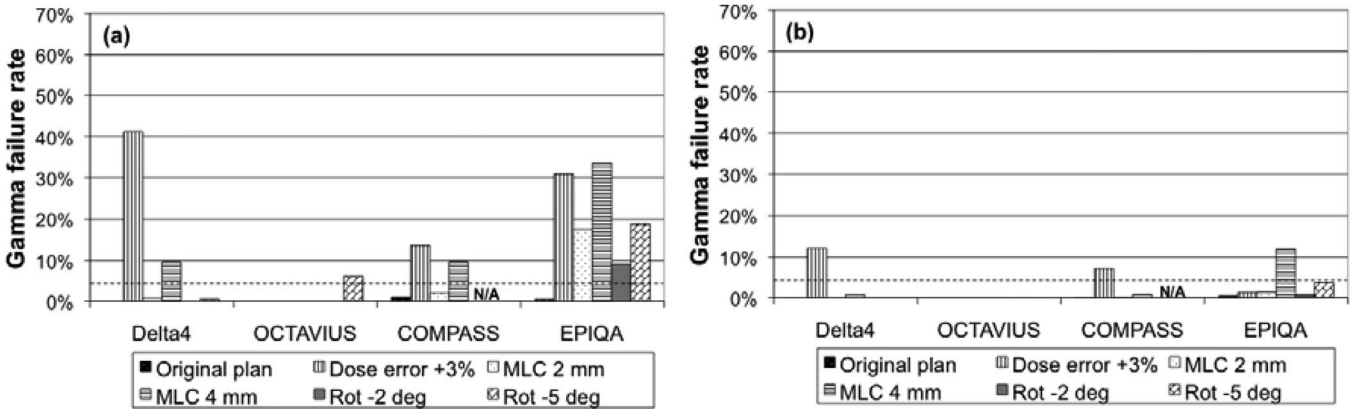

FIG. 4. The results for the second brain case, in (a) the gamma criteria of 2%/2 mm was used and in (b) 3%/3 mm was used. The dashed line represents a failure rate of 5%.

The approach we used in this work was to introduce errors that can occur based on machine malfunctions or calibration errors. Therefore, the errors were selected first and then the dosimetric error resulting from the error was subsequently quantified. Indeed, some of the errors proved to introduce rather limited dosimetric effects in a patient geometry. This is rather instructive too, although it is not the focus of the present work. We were curious how this would be manifested in the QA system measurements, given the marked deviation from a patient geometry in the detector geometry of some of the systems. In this work, we show that some of the systems pick up and highlight minor errors that are probably less clinically relevant. In addition to this fact, due to uncertainties in the delivery, plans will be both off on the mean and with a spread around this mean in various parts of the geometry. This effect is plan and system dependent. Therefore, the plans with +3% dose will be more likely to be off on the gamma analysis but are not necessarily always detected by all systems in a gamma analysis.

Since there might be systematic differences between the performances of the machine with that at the time of commission of the machine for the TPS, it would be preferable to compare the measurement of the original plan with the measurement of the plans with errors. In that case the influence of the TPS would be removed. Unfortunately, all systems did not allow for a comparison of two measurements and, therefore, more plans could potentially be seen as failed given the fact that calculated plans were compared with the measurements. Also, the calibration of the patient QA system may introduce systematic errors. In this study, the patient QA system was calibrated at the time of measurement in order to correct for (long-term) variations in machine output in terms of dose per monitor unit, assuming that this parameter is checked using other systems (which is the case in our institution).

As described in Sec. II we chose to use different types of plans to see if there were any differences in the result. In testing more cases, we not only expect some varieties within the same group of diseases as seen between the brain cases but also some prominent differences between the different disease groups. For example, from the data presented herein we expect the rotational error to be more prominent in more complex plans with nonspherical targets (such as the head and neck plan) compared to the more spherical brain cases where the rotational error did not have the same impact.

The OCTAVIUS® system detected the fewest errors, 3 of 20 when using the 3%/3 mm criteria and 8 of 20 when using 2%/2 mm criteria. This was possibly a result of the procedure involving CT scanning of the phantom with the

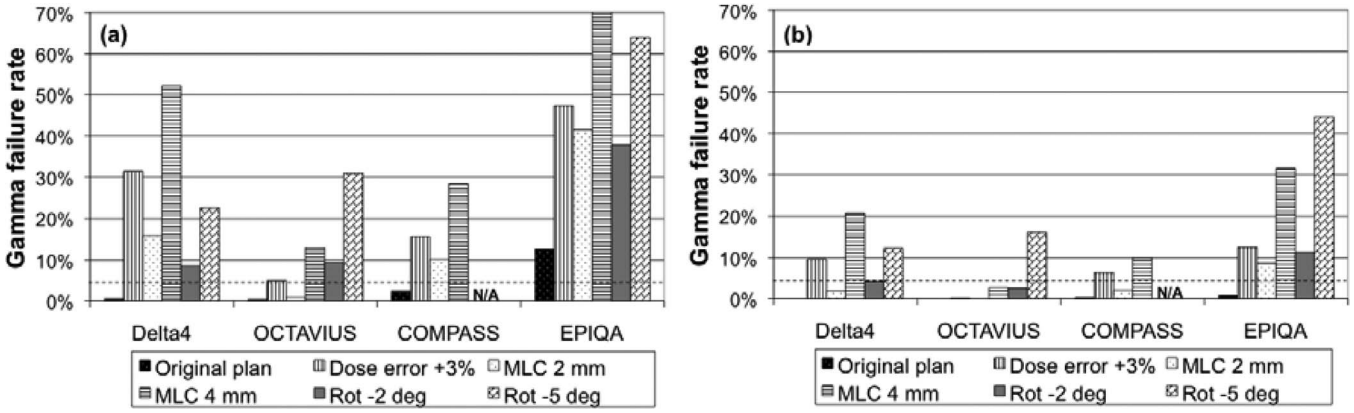

FIG. 5. The results for the head and neck case, in (a) the gamma criteria of 2%/2 mm was used and in (b) 3%/3 mm was used. The dashed line represents a failure rate of 5%.

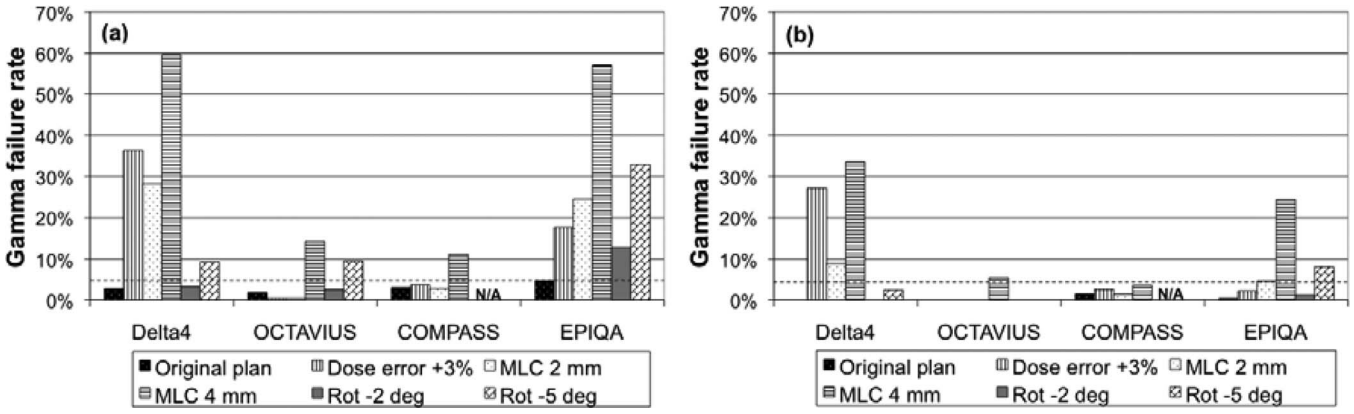

FIG. 6. The results for the prostate case, in (a) the gamma criteria of 2%/2 mm was used and in (b) 3%/3 mm was used. The dashed line represents a failure rate of 5%.

ionization chamber matrix inserted (as suggested by the manufacturer). This resulted in a "jagged" profile of Hounsfield units, producing, in turn, a jagged dose profile. In principle, a jagged profile could tend to pass most points in a gamma analysis given that a matching dose value is likely to be found. Better results may therefore be obtained when optimizing the methodology, which was beyond the scope of this study.

For the Delta$^{4®}$ the 3% dose error and the 4 mm MLC error had the highest gamma failure rate and the collimator rotation error had the lowest gamma failure rate. In total Delta$^{4®}$ detected 9 of 20 errors when using the 3%/3 mm criteria and 15 of 20 errors when using the 2%/2 mm criteria and all of the original plans had a failure rate less than 95% for both criteria. In addition to the gamma evaluation it was also possible to import the structures and compare the DVH as shown in Fig. 7, where the DVH for the prostate case with 3% dose error is shown. This can be an extra help when evaluating the results, although one must remember that the DVH analysis are for structures transferred to the phantom (both the measured ones and the calculated ones), and that the actual anatomy of the patient is not included in the evaluation.

For the COMPASS system an isodose curve of 10% was used for the gamma evaluation, to match the other systems, which have a lower dose cutoff at 10%. But the intention with the COMPASS system is to look at structure by structure and when doing so the gamma failure rate was much higher for some of the structures and more of the errors can be considered detected. For example, in the prostate case the failure rate for the 10% isodose structure (using 2%/2 mm criteria), for the 3% dose error and the 4 mm MLC error, was 4% and 11%, respectively. Whereas the failure rate for the PTV structure was 83% and 88%, respectively, which was also clearly visible when looking at the DVH, see Fig. 8. When using 3%/3 mm criteria the COMPASS system detected 5 of the errors with the 10% isodose structure and detected 8 of the errors with the PTV structure. With 2%/2 mm criteria the number of detected errors was 8 and 11, respectively.

Epiqa^{TM} was the system that detected the most errors, 11 errors with the 3%/3 mm criteria and all of the errors with 2%/2 mm criteria, but with a high failure rate for the original H&N plan (13% for 2%/2 mm criteria). When looking at the Epiqa^{TM} results one must also remember that the area that is

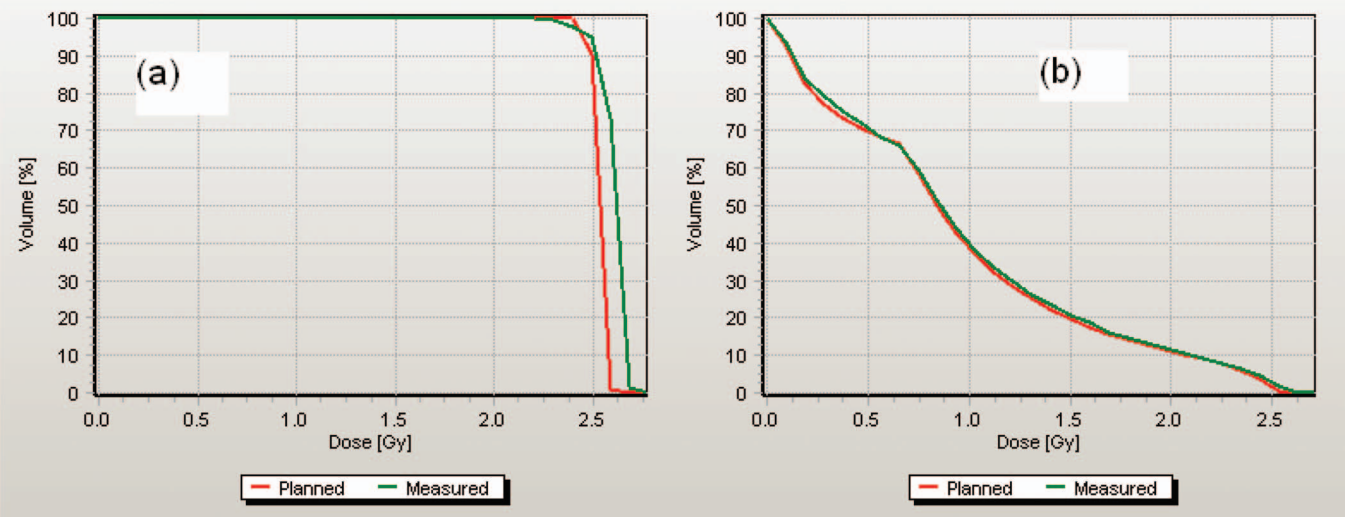

FIG. 7. DVH from the Delta$^{4®}$ system for the prostate case with the 3% dose error. The left graph (a) is the planned and measured DVH for the prostate and the right graph (b) is the planned and measured DVH for the rectum.

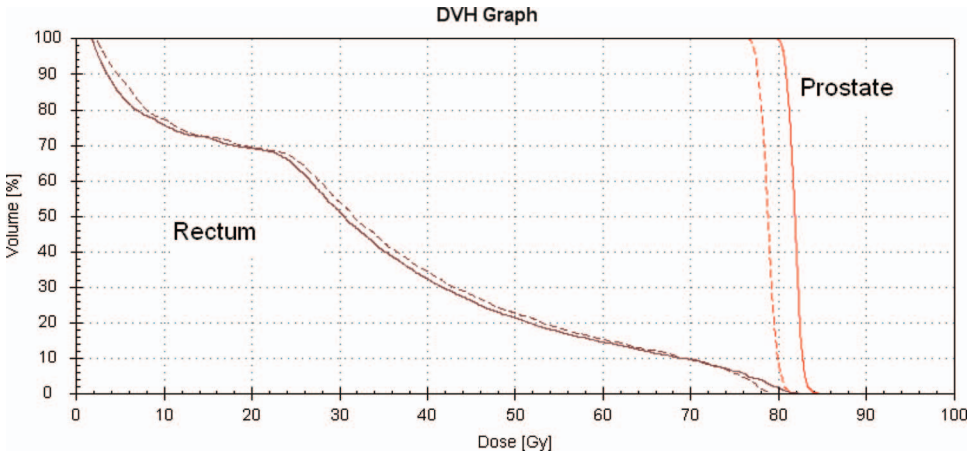

FIG. 8. DVH from the COMPASS system for the prostate case with 3% dose error. The dashed lines are the planned dose and the solid lines are the measured ones.

taken into account in the analysis includes the total of the field size bounded by the outermost MLC limits, while the others include every dose point from 10% of maximum dose and above. Our intentions have been to do a fair comparison, but because the systems inherent differences and that the different software did not allow the same analysis methods it is difficult to say that any system is better than the other.

The results in this study show that the 3%/3 mm criteria might not be sensitive enough to detect all errors of the magnitude that has been introduced in this study. One might, therefore, be tempted to suggest that 2%/2 mm criteria is better, but if lowering the criteria the risk of detecting errors that are not of importance will increase. When performing patient specific QA each clinic, therefore, has to think about why they are doing the measurements; what kind of errors they are looking for and want to detect with the measurement. Also, what impact will the error have on the total treatment. Some of the errors introduced in this study might not have a large impact on the total treatment while others have. For example, a 3% increase in MU will always render an increased dose, whereas a collimator rotation might not always have an important impact on the treatment. We, therefore, claim that it is important that each clinic carefully considers which action level is suitable for their methods and quality assurance system.

V. CONCLUSION

In this study four rotational IMRT plans have been measured with four different commercial QA systems designed for rotational therapy. Measurements of plans with intentional errors such as increased dose to target, widening of the MLC bank, and rotation of the collimator have been made to test the various systems sensitivity.

In this set of experiments, the systems tended toward identifying errors with varying degree of sensitivity, suggesting that there might be marginal benefit in the use of multiple systems. The possible exception was the OCTAVIUS® and COMPASS systems, where the OCTAVIUS® system detected five plans out of eight with collimator rotation errors.

When designing the complete QA program, it should be taken into account that the patient QA system was designed to test certain parts of the radiation therapy delivery, and that this varies from system to system. Overall, we find that all the systems perform well in terms of detecting errors. To this end, the present study can form a basis for evaluating and altering the supplementary quality assurance procedures of the radiation therapy. We conclude that the sensitivity to the introduced errors depends on the plan and the various systems can detect various errors, and some of the errors cannot be found with these systems. There were also poor correlations between the gamma evaluation pass rates of the QA procedures and the deviations observed in the DVH. Furthermore, the calibration of the systems appears to be important and the measuring procedure or calibration as described by the manufacturer might benefit from improvements in some cases.

ACKNOWLEDGMENTS

Thanks to all the manufacturers for the support and valuable discussions, and to Sun Nuclear, IBA, and PTW for loan of equipment.

[a]Author to whom correspondence should be addressed. Electronic mail: per.munck@rh.regionh.dk; Fax +45 3545 9593.

[1]M. Alber *et al.*, *ESTRO Booklet No. 9 Guidelines for the Verification of IMRT*, edited by G. Mijnheer (ESTRO, Brussel, 2008).
[2]C. X. Yu and G. Tang, "Intensity-modulated arc therapy: Principles, technologies and clinical implementation," Phys. Med. Biol. **56**, R31–R54 (2011).
[3]C. C. Ling, P. Zhang, Y. Archambault, J. Bocanek, G. Tang, and T. Losasso, "Commissioning and quality assurance of RapidArc radiotherapy delivery system," Int. J. Radiat. Oncol., Biol., Phys. **72**, 575–581 (2008).
[4]X. A. Li, L. Ma, S. A. Naqvi, R. Shih, and C. X. Yu, "Monte Carlo dose verification for intensity-modulated arc therapy," Phys. Med. Biol. **46**, 2269–2282 (2001).
[5]K. Bush, R. Townson, and S. Zavgorodni, "Monte Carlo simulation of RapidArc radiotherapy delivery," Phys. Med. Biol. **53**, N359–N370 (2008).
[6]E. Schreibmann, A. Dhabaan, E. Elder, and T. Fox, "Patient-specific quality assurance method for VMAT treatment delivery," Med. Phys. **36**, 4530–4535 (2009).
[7]T. Teke, A. M. Bergman, W. Kwa, B. Gill, C. Duzenli, and I. A. Popescu, "Monte Carlo based, patient-specific RapidArc QA using Linac log files," Med. Phys. **37**, 116–123 (2010).

[8]M. Iori, E. Cagni, A. E. Nahum, and G. Borasi, "IMAT-SIM: A new method for the clinical dosimetry of intensity modulated arc therapy (IMAT)," Med. Phys. **34**, 2759–2773 (2007).
[9]M. Iori, E. Cagni, M. Palusco, P. Munro, and A. E. Nahum, "Dosimetric verification of IMAT delivery with a conventional EPID system and a commercial portal dose image prediction tool," Med. Phys. **37**, 377–390 (2010).
[10]K. Vergote, Y. De Deene, W. Duthoy, W. De Gersem, W. De Neve, E. Achten, and C. De Wagter, "Validation and application of polymer gel dosimetry for the dose verification of an intensity-modulated arc therapy (IMAT) treatment," Phys. Med. Biol. **49**, 287–305 (2004).
[11]S. Ceberg, I. Gagne, H. Gustafsson, J. B. Scherman, S. S. Korreman, F. Kjær-Kristoffersen, M. Hilts, and S. Å. J. Bäck, "RapidArc treatment verification in 3D using polymer gel dosimetry and Monte Carlo simulation," Phys. Med. Biol. **55**, 4885–4898 (2010).
[12]S. Korreman, J. Medin, and F. Kjær-Kristoffersen, "Dosimetric verification of RapidArc treatment delivery," Acta Oncol. **48**, 185–191 (2009).
[13]A. Haga, K. Nakagawa, K. Shiraishi, S. Itoh, A. Terahara, H. Yamashita, K. Ohtomo, S. Saegusa, T. Imae, K. Yoda, and R. Pellegrini, "Quality assurance of volumetric modulated arc therapy using Elekta Synergy," Acta Oncol. **48**, 1193–1197 (2009).
[14]I. Iftimia, E. T. Cirino, L. Xiong, and H. W. Mower, "Quality assurance methodology for Varian RapidArc treatment plans," J. Appl. Clin. Med. Phys. **11**, 130–143 (2010).
[15]V. Chandraraj, S. Stathakis, R. Manickam, C. Esquivel, S. S. Supe, and N. Papanikolaou, "Comparison of four commercial devices for RapidArc and sliding window IMRT QA," J. Appl. Clin. Med. Phys. **12**, 338–349 (2011).
[16]L. Masi, F. Casamassima, R. Doro, and P. Francescon, "Quality assurance of volumetric modulated arc therapy: Evaluation and comparison of different dosimetric systems," Med. Phys. **38**, 612–621 (2011).
[17]G. J. Kutcher, L. Coia, M. Gillin, W. F. Hanson, S. Leibel, R. J. Morton, J. R. Palta, J. A. Purdy, L. E. Reinstein, G. K. Svensson, M. Weller, and L. Wingfield, "Comprehensive QA for radiation oncology: Report of AAPM Radiation Therapy Committee Task Group 40," Med. Phys. **21**, 581–618 (1994).
[18]V. Feygelman, G. Zhang, C. Stevens, and B. E. Nelms, "Evaluation of a new VMAT QA device, or the "X" and "O" array geometries," J. Appl. Clin. Med. Phys. **12**, 146–168 (2011).
[19]J. L. Bedford, Y. K. Lee, P. Wai, C. P. South, and A. P. Warrington, "Evaluation of the Delta4 phantom for IMRT and VMAT verification," Phys. Med. Biol. **54**, N167–N176 (2009).
[20]E. Spezi, A. Angelini, F. Romani, and A. Ferri "Characterization of a 2D ion chamber array for the verification of radiotherapy treatments," Phys. Med. Biol. **50**, 3361–3373 (2005).
[21]A. Van Esch, C. Clermont, M. Devillers, and M. Iori, "On-line quality assurance of rotational radiotherapy treatment delivery by means of a 2D ion chamber array and the Octavius phantom," Med. Phys. **34**, 3825–3837 (2007).
[22]R. Boggula, F. Lorenz, L. Mueller, M. Birkner, H. Wertz, F. Stieler, V. Steil, F. Lohr, and F. Wenz, "Experimental validation of a commercial 3D dose verification system for intensity-modulated arc therapies," Phys. Med. Biol. **55**, 5619–5633 (2010).
[23]G. Nicolini, E. Vanetti, A. Clivio, A. Fogliata, S. Korreman, J. Bocanek, and L. Cozzi, "The GLAaS algorithm for portal dosimetry and quality assurance of RapidArc, an intensity modulated rotational therapy," Radiat. Oncol. **3** (2008).
[24]A. Fogliata, A. Clivio, P. Fenoglietto, J. Hrbaeck, S. Kloeck, P. Lattauda, P. Mancosu, G. Nicolini, E. Parietti, G. Urso, E. Vanetti, and L. Cozzi, "Quality assurance of RapidArc in clinical practice using portal dosimetry," Br. J. Radiol. **84**, 534–545 (2011).
[25]D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," Med. Phys. **25**, 656–661 (1998).
[26]B. E. Nelms, H. Zhen, and W. A. Tomé, "Per-Beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors," Med. Phys. **38**, 1037–1044 (2011).
[27]H. Zhen, B. E. Nelms, and W. A. Tomé, "Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA," Med. Phys. **38**, 5477–5489 (2011).