


# Sensitivity of Three Patient-Specific Quality Assurance Systems to MLC Aperture Errors With Volumetric Modulated Arc Therapy

Technology in Cancer Research & Treatment  
Volume 21: 1-13  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/15330338221114499  
journals.sagepub.com/home/tct  


Jinyou Hu, MS<sup>1,2</sup> , Shaoxian Gu, MS<sup>1</sup>, Ningyu Wang, MS<sup>1</sup>, Fengjie Cui, MS<sup>1</sup>, Shengyuan Zhang, MS<sup>1</sup>, Chuou Yin, MS<sup>1</sup>, Yunzhu Cai, MS<sup>1</sup>, Chengjun Gou, PhD<sup>1</sup>, Lian Zou, PhD<sup>2</sup>, and Zhangwen Wu, PhD<sup>1</sup>

## Abstract

**Purpose:** To compare the sensitivity of ArcCHECK (AC), portal dosimetry (PD), and an in-house logfile-based system (LF) to multileaf collimators (MLC) aperture errors and the ability to identify these errors. **Methods and Materials:** For 12 retrospective original head and neck volumetric modulated arc therapy (VMAT) plans, MLC aperture errors of  $\pm 0.4\text{mm}$ ,  $\pm 1.2\text{mm}$ ,  $\pm 2\text{mm}$ , and  $\pm 3\text{mm}$  were introduced for each plan, resulting in 96 plans with errors. AC, PD, and LF were used for the gamma evaluation at 3%/3mm, 3%/2mm, and 2%/2mm criteria. Gradient analysis was used to evaluate the sensitivity to MLC aperture errors. The area under the curve (AUC) obtained from the receiver operating characteristic (ROC) curve was used to evaluate the ability to identify MLC aperture errors and dose errors, and the optimal cut-off value to identify the error was obtained. **Results:** The gamma pass rate (%GP) of LF had the smallest descent gradient as the MLC error increases in any case. The descent gradient of PD was larger than AC, except for the case at the 2%/2mm criteria. For the 3%/3mm criteria, the MLC aperture errors that can be perfectly identified by AC, PD, and LF were  $\pm 3\text{mm}$ ,  $\pm 2\text{mm}$ , and  $\pm 1.2\text{mm}$ , respectively, and the average percent dose error (%DEs) of dose metrics in targets that can be perfectly identified were 4% to 5%, 3% to 4%, and 2% to 3%, respectively. For the 3%/2mm criteria, the errors that AC, PD, and LF can perfectly identify were the same as the 3%/3mm criteria. For the 2%/2mm criteria, AC can perfectly identify the MLC error of  $\pm 2\text{mm}$  and the %DE of 3% to 4%. PD and LF can identify the MLC error of  $\pm 1.2\text{mm}$  and the %DE of 2% to 3%. **Conclusion:** Different patient-specific quality assurance (PSQA) systems have different sensitivity and recognition abilities to MLC aperture errors. Institutions should formulate their own customized %GP limits based on their PSQA process through ROC or other methods.

## Keywords

ArcCHECK, portal dosimetry, logfile, MLC aperture errors, VMAT

## Abbreviations

AC, ArcCHECK; AUC, area under the curve; AAPM, American Association of Physicist in Medicine; AAA, Anisotropic Analytical Algorithm; BR\_PRV, planning organ at risk volume of brainstem; CT, computerized tomography; CTVI, high-risk clinical target

<sup>1</sup> Key Laboratory of Radiation Physics and Technology of the Ministry of Education, Institute of Nuclear Science and Technology, Sichuan University, Chengdu, China

<sup>2</sup> Cancer Center, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, China

## Corresponding Authors:

Zhangwen Wu, PhD, Key Laboratory of Radiation Physics and Technology of the Ministry of Education, Institute of Nuclear Science and Technology, Sichuan University, No. 29, Wangjiang Road, Chengdu 610064, China.

Email: wuzhangwen@scu.edu.cn

Lian Zou, PhD, Cancer Center, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, No. 32, W. Sec 2, 1st Ring Rd, Chengdu 610072, China.

Email: zoulilian@med.uestc.edu.cn



volume; CTV2, low-risk and preventive radiation volume; DVH, dose-volume histogram; %DE, dose percentage error; DD, dose difference; DTA, distance to agreement; EPID, electronic portal imaging device; DLG, dosimetric leaf gap; %GP, gamma pass rate; GTVnx, gross tumor volume of the nasopharynx; GTVnd, involved lymph nodes; gEUD, generalized equivalent uniform dose; IMRT, intensity-modulated radiation therapy; LF, logfile-based system; LINAC, linear accelerator; MLC, multileaf collimators; MU, monitor unit; OAR, organ at risk; PSQA, patient-specific quality assurance; PD, portal dosimetry; PO, photon optimizer; PDIP, Portal Dose Image Prediction; PGTVnx, planning target of gross tumor volume of the nasopharynx; PGTVnd, planning target of involved lymph nodes; PTVI, planning target of high-risk clinical target volume; PTV2, planning target of low-risk and preventive radiation volume; QA, quality assurance; ROC, receiver operating characteristics; SD, standard deviation; SP\_PRV, planning organ at risk volume of spinal cord; TG, task group; TG218, Task Group No. 218; TPS, treatment planning system; VMAT, volumetric modulated arc therapy.

Received: January 31, 2022; Revised: June 14, 2022; Accepted: July 1, 2022.

## Introduction

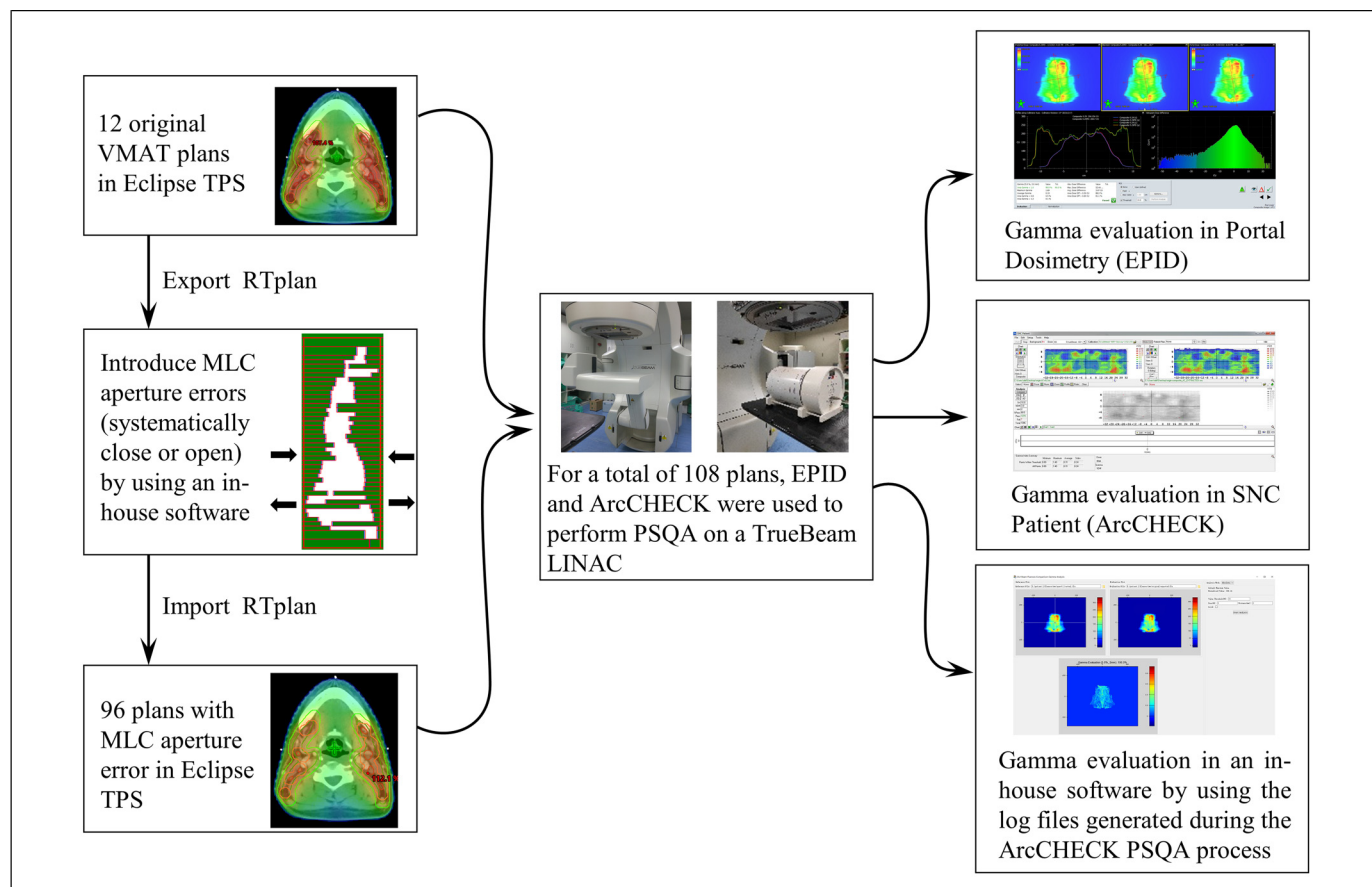
To ensure the safety and accuracy of radiotherapy, pre-treatment patient-specific quality assurance (PSQA) is a very important process to verify the consistency between the calculated and delivered doses.<sup>1</sup> Volumetric modulated arc therapy (VMAT) is becoming the preferred treatment technology for radiotherapy due to significantly improved treatment efficiency while maintaining plan quality.<sup>2-5</sup> However, there may be different individual delivery errors since VMAT has more degrees of freedom, such as position errors of multileaf collimators (MLC), gantry, collimator, and changes in dose rate and output. This demands higher requirements and challenges to meet PSQA guidelines. Previous studies showed that changes in the dose-volume histogram (DVH) caused by the MLC aperture opening/closing errors are larger compared to other errors.<sup>6-9</sup> These MLC errors have a larger impact on dose distribution, which could compromise the patient's clinical outcome. MLC position accuracy is essential for a complex VMAT plan delivery. Therefore, PSQA should focus more on the detection of leaf gap errors.

The primary method to determine the consistency between 2 dose distributions is to use gamma evaluation, as proposed by Low et al.<sup>10,11</sup> This method combines the dose difference (DD) and distance to agreement (DTA) criteria to calculate the gamma index of each point. The gamma pass rate (%GP) is the ratio of points passing the gamma evaluation (gamma index < 1) to all points. Different devices can be used for VMAT PSQA in clinical practice. The devices that are based on phantom measurement are the most commonly used, such as ArcCHECK (AC) (Sun Nuclear Corporation), Delta4 (ScandiDos), and Octavius (PTW). Quality assurance (QA) based on phantom measurement usually requires a complex phantom setup and additional machine delivery time, which is time-consuming and labor-intensive. Another commonly used device is the electronic portal imaging device (EPID), using software such as portal dosimetry (PD) (Varian Medical System) or Perfraction (Sun Nuclear Corporation) for analysis of the measurement. The EPID is usually installed on the medical linear accelerator (LINAC). Therefore, this type of device, which does not require additional setup, is more efficient than the device based on phantom measurement.

In recent years, a more time and labor-efficient QA method, based on logfiles, has attracted the interest of researchers. Almost all of the delivered parameters of the LINAC can be conveniently entered in the logfile to enable machine performance QA. Some studies have validated the accuracy of the information in the logfile, which supports the use of logfiles for PSQA.<sup>12,13</sup> Logfile-based QA does not require a phantom for measurement. Following delivery, the data are automatically recorded and can be automatically evaluated by some QA systems, thereby avoiding setup errors, reducing labor costs, and significantly improving PSQA efficiency. However, it has been reported that the parameters recorded in the logfile are not always correct.<sup>14</sup> Another contentious factor of logfile-based QA is its independence. It is not measured by a third-party system but from the LINAC itself. Therefore, whether logfile-based QA can replace measurement-based QA remains a controversial topic among physicists.

Although several systems can be used for PSQA and gamma evaluation, the results in equivalent conditions are not the same and this difference in results is not widely understood. Therefore, users are unsure which results can be trusted, particularly when one plan fails validation on one system but passes or is approved on another system. Moreover, there is no uniform standard for the method and result analysis of plan verification among radiotherapy institutions. The accuracy and comparability of different systems and their detection ability of errors are issues that have attracted much attention. To promote PSQA standardization, the American Association of Physicists in Medicine (AAPM) Task Group No. 218 (TG218) conducted relevant investigations and studies and proposed the methods as well as standards of gamma evaluation, which recommended the universal tolerance limits and action limits of %GP to be 95% and 90% for gamma analysis using global normalization in absolute dose with 3%/2mm (DD/DTA) criteria and a 10% dose threshold.<sup>15</sup> However, using the same gamma analysis criteria and limits for all PSQA systems may be inappropriate because the %GP of different systems may vary widely.

The receiver operating characteristic curve (ROC) is a common method used in clinical studies to compare the ability of different tests to distinguish between negative and positive. ROC has been used to analyze the ability to detect radiotherapy errors in several studies.<sup>16-19</sup> The ROC curve is the line connecting each point with



**Figure 1.** Flowchart of the research process.

a false positive rate (1-Specificity) as the abscissa and the true positive rate (Sensitivity) as the ordinate under different thresholds of a test. The value of the area under the curve (AUC) is between 0.5 and 1. The higher the value, the better the test's ability to distinguish between positive and negative. The corresponding optimal cut-off value can be calculated through ROC analysis.

This study used AC, PD, and an in-house logfile-based system (LF) to perform PSQA on the retrospective head and neck VMAT plans and plans generated by introducing MLC aperture open/close errors to these plans. We investigated the differences between the systems and evaluated their sensitivity to MLC aperture errors based on gamma analysis results. The impact of MLC aperture errors on the dose distribution of the targets and organs at risk (OARs) and the correlation between the gamma results and the dose error were analyzed. At the same time, ROC was used to obtain the AUC and the optimal cut-off value for detecting a certain MLC aperture or dose error for different PSQA systems under various gamma analysis criteria.

## Materials and Methods

### Patient and Study Plan

The research process is shown as a flowchart in Figure 1. From the clinical database, we obtained 12 retrospective head and neck

VMAT plans that were treated. All the plans had 2 full arcs, of which the gantry of one rotated clockwise from  $181^\circ$  to  $179^\circ$ , and the collimator was set around  $10^\circ$ . The other arc rotated back counterclockwise, and the collimator was set around  $350^\circ$ . The dose was calculated using the AAA (Anisotropic Analytical Algorithm) algorithm with a 2.5mm grid in the Eclipse (version 13.6, Varian Medical System) treatment planning system (TPS) at 6MV of photon energy. All the plans were delivered on a TrueBeam LINAC (Varian Medical System) equipped with Millennium 120-leaf MLC. One oncologist contoured the targets and OARs based on the planning CT. The planning targets named planning target of gross tumor volume of the nasopharynx (PGTVnx), planning target of involved lymph nodes (PGTVnd), planning target of high-risk clinical target volume (PTV1), and planning target of low-risk and preventive radiation volume (PTV2) were generated by adding a margin of about 3 to 5mm to gross tumor volume (GTVnx), involved lymph nodes (GTVnd), high-risk clinical target volume (CTV1), and low-risk and preventive radiation volume (CTV2), respectively. Similarly, the planning organ at risk volume (PRV), named planning organ at risk volume of brainstem (BR\_PRV) and planning organ at risk volume of spinal cord (SP\_PRV), were generated from the brainstem and spinal cord, respectively. The prescribed doses were 69.96Gy/33f for PGTVnx and PGTVnd, 62.7Gy/33f for PTV1, and 54.45Gy/33f for PTV2, respectively.

These 12 plans were regarded as the original plans. Next, the RTplan files of the original plans were exported to an in-house software to edit the position of each leaf in the bank on both sides under the jaw aperture at every control point to move in the opposite direction by 0.2mm, 0.6mm, 1mm, and 1.5mm, meaning that the MLC aperture systematically opened or closed 0.4mm, 1.2mm, 2mm, and 3mm, respectively. Each original plan formulated 8 plans that contained MLC errors, meaning that there were 96 plans with MLC errors. These plans were then imported into the Eclipse TPS for dose calculation. The absolute dose percentage error (%DE) of the MLC error plan and original plan were compared using the following dose metrics:  $D_{98\%}$ ,  $D_{2\%}$  (dose received by 98% and 2% of the volume), and generalized equivalent uniform dose (gEUD) for the planning target,  $D_{\max}$  (maximum dose), and gEUD for BR\_PRV and SP\_PRV, and gEUD for the left and right parotid. The gEUD is calculated by the following formula:

$$\left[ gEUD = \left( \sum_i v_i D_i^a \right)^{1/a} \right] \quad (1)$$

where  $v_i$  is the fractional organ volume receiving a dose  $D_i$  and  $a$  is a tissue-specific parameter that describes the volume effect.<sup>20,21</sup> Based on AAPM TG166, the  $a$  value of the target, BR\_PRV, SP\_PRV, and parotid was  $-10$ ,  $16$ ,  $20$ , and  $1$ , respectively.<sup>21</sup>

The absolute %DE is calculated as follows:

$$\left[ \%DE = \frac{\text{MLC error plan dose} - \text{Original plan dose}}{\text{Original plan dose}} \times 100\% \right]. \quad (2)$$

The gradient analysis was used to evaluate the %DE changes in per unit error, which was obtained by linear regression between %DE and MLC aperture errors.

We used AC and EPID to perform PSQA on TrueBeam LINAC for 108 plans. The logfiles generated during the AC measurement were used for logfile-based PSQA. All measurements were completed within 2 consecutive days to minimize daily output variation. All PD and AC measurements were completed on the first and second days, respectively. The machine QA of LINAC including absolute dose calibration using DOSE1 (IBA Dosimetry) was performed each day prior to PSQA systems calibration and measurement. The dose and array response calibrations of AC and EPID were implemented before the measurement.

### ArcCHECK

The AC, Model 1220, is a three-dimensional beam dosimetry QA system, with 1386 diode detectors of  $0.8 \times 0.8$  mm size and 10mm sensor spacing, arranged in a spiral pattern on a cylindrical water-equivalent phantom with 21cm array diameter and length.

The original plan was calculated on the AC phantom using the same algorithm and grid. The dose distribution was

exported to SNC Patient software (version 6.2.3, Sun Nuclear Corporation) for gamma evaluation. The AC phantom was then irradiated to measure the dose distribution of each plan. All the measured distributions were compared with the calculated distribution of the original plan.

### Portal Dosimetry

The PD (version 13.6) is a system to perform gamma evaluation based on EPID measurement data. The TrueBeam LINAC has an amorphous silicon EPID (Portal vision aS1000) with  $1024 \times 768$  pixels and a measuring area of  $40\text{cm} \times 30$  cm. PD verification plan was created for all the plans using the Portal Dose Image Prediction (PDIP) algorithm. The data measured by EPID are automatically recorded into the corresponding verification plan once these verification plans are delivered on the TrueBeam LINAC.

### Logfile-Based System

TrueBeam synchronized the planned and actual delivery with an update rate of 20ms. The information recorded in the logfile includes the MU, position of MLC and jaw, angle of gantry and collimator, etc. In this study, the logfiles generated during the AC PSQA process were used for evaluation. In-house software was created using Python script based on Pylinac v2.5.1<sup>22</sup> and Flashgamma<sup>23,24</sup> to perform the logfile-based evaluation. The leaf position, jaw position, and MU were extracted from the logfile to generate the actual and expected fluence map with a 1mm resolution. Gamma evaluation was performed between the delivered actual fluence map of each plan and the expected fluence map of the original plan.

### Gamma Evaluation

For all 3 systems, the dose distributions of the 2 arcs of each plan were computed into one. Next, gamma evaluation was performed between the measured distribution of each plan and the computed distribution of the original plan. In this study, 3 different criteria of 3%/3mm, 3%/2mm, and 2%/2mm, with the recommended settings of TG218, were used for gamma evaluation.

### Data Analysis and Statistics

Sensitivity was expressed by the gradient and the area under the ROC curve (AUC). The gradient was the slope obtained by linear regression between %GP and MLC aperture errors, which equals the change in the value of %GP per unit error. The larger the absolute value of the gradient, the larger the response of the corresponding system to the errors of the MLC aperture.

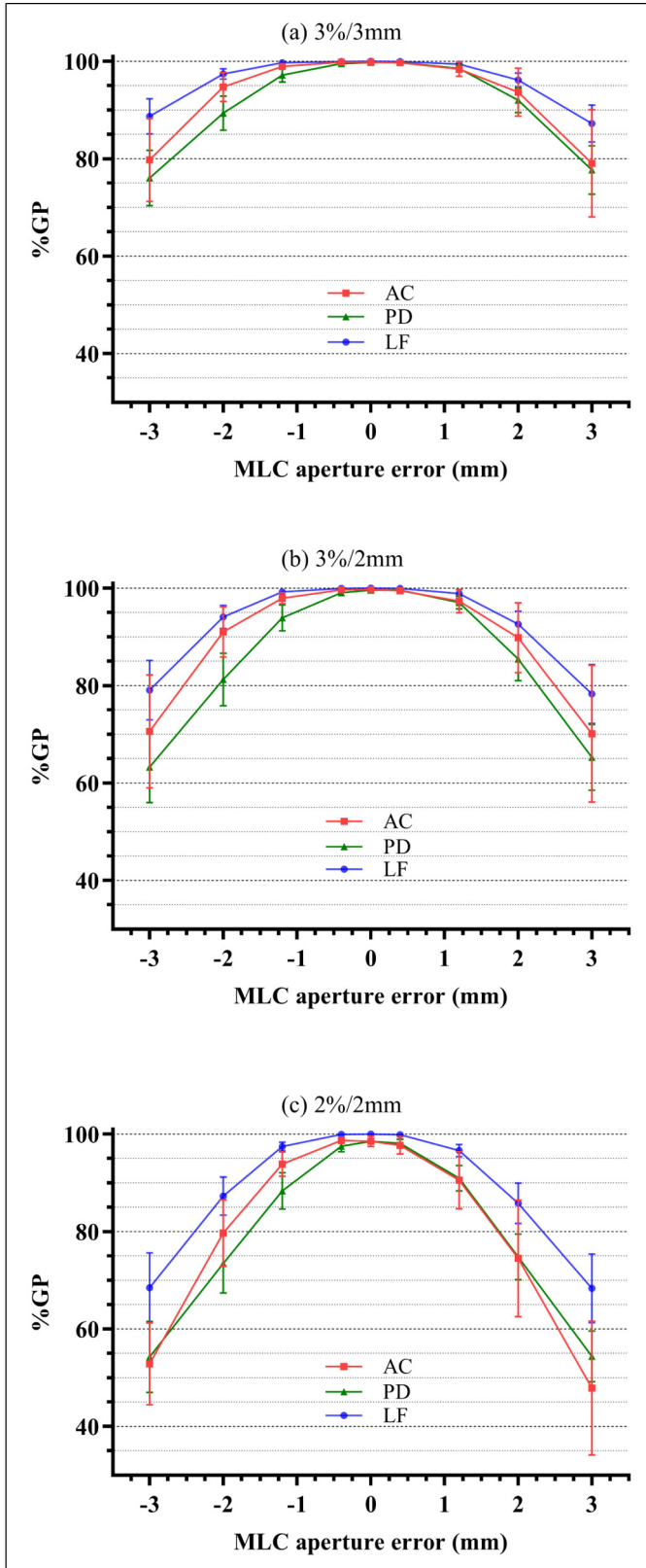
The correlation between the %GP at different DD/DTA criteria and the absolute %DE including all dose metrics of all targets or all OARs was analyzed by linear fitting. The average of the absolute %DE of all dose metrics for all

**Table 1.** Percentage Dose Error (%DE), Mean  $\pm$  Standard Deviation [SD] of Dose Metrics Between the Plans With Multileaf Collimators (MLC) Aperture Errors and the Original Plan and the Equation of Linear Fitting Between %DE and MLC Aperture Error.

Structure	Dose metrics	%DE <sup>a</sup> caused by MLC aperture error (mean $\pm$ SD, %)											Equation	R <sup>2</sup>
		-0.4mm	-1.2mm	-2mm	-3mm	0.4mm	1.2mm	2mm	3mm					
PGTVnx	D <sub>98%</sub>	-1.26 $\pm$ 0.40	-3.82 $\pm$ 1.13	-6.44 $\pm$ 1.77	-9.79 $\pm$ 2.47	1.25 $\pm$ 0.42	3.68 $\pm$ 1.31	5.99 $\pm$ 2.23	8.76 $\pm$ 3.42	Y = 3.1X - 0.18	0.91			
	D <sub>2%</sub> gEUD	-0.96 $\pm$ 0.20	-2.75 $\pm$ 0.56	-4.45 $\pm$ 0.90	-6.49 $\pm$ 1.32	1.01 $\pm$ 0.21	3.14 $\pm$ 0.6	5.41 $\pm$ 0.95	8.35 $\pm$ 1.37	Y = 2.47X + 0.36	0.96			
PGTVnd	D <sub>98%</sub>	-1.06 $\pm$ 0.25	-3.20 $\pm$ 0.74	-5.38 $\pm$ 1.21	-8.12 $\pm$ 1.78	1.07 $\pm$ 0.27	3.17 $\pm$ 0.79	5.23 $\pm$ 1.29	7.73 $\pm$ 1.90	Y = 2.65X - 0.06	0.95			
	D <sub>2%</sub> gEUD	-0.78 $\pm$ 0.18	-2.48 $\pm$ 0.57	-4.42 $\pm$ 1.03	-7.04 $\pm$ 1.64	0.72 $\pm$ 0.16	2.04 $\pm$ 0.45	3.24 $\pm$ 0.72	4.60 $\pm$ 1.02	Y = 1.93X - 0.46	0.94			
PTV1	D <sub>98%</sub>	-0.62 $\pm$ 0.13	-1.80 $\pm$ 0.38	-2.95 $\pm$ 0.62	-4.32 $\pm$ 0.89	0.65 $\pm$ 0.13	2.07 $\pm$ 0.45	3.57 $\pm$ 0.86	5.50 $\pm$ 1.40	Y = 1.63X + 0.23	0.95			
	D <sub>2%</sub> gEUD	-0.71 $\pm$ 0.27	-2.24 $\pm$ 0.91	-3.86 $\pm$ 1.64	-6.01 $\pm$ 2.78	0.72 $\pm$ 0.23	2.07 $\pm$ 0.61	3.38 $\pm$ 0.94	4.94 $\pm$ 1.30	Y = 1.82X - 0.19	0.88			
PTV2	D <sub>98%</sub>	-1.30 $\pm$ 0.30	-3.95 $\pm$ 0.87	-6.69 $\pm$ 1.37	-10.13 $\pm$ 1.93	1.29 $\pm$ 0.31	3.82 $\pm$ 0.96	6.25 $\pm$ 1.63	9.13 $\pm$ 2.47	Y = 3.22X - 0.18	0.95			
	D <sub>2%</sub> gEUD	-0.94 $\pm$ 0.18	-2.71 $\pm$ 0.51	-4.41 $\pm$ 0.82	-6.42 $\pm$ 1.21	0.98 $\pm$ 0.19	3.04 $\pm$ 0.58	5.20 $\pm$ 0.97	7.98 $\pm$ 1.45	Y = 2.4X + 0.30	0.96			
BR_PRV	D <sub>98%</sub>	-1.10 $\pm$ 0.19	-3.33 $\pm$ 0.57	-5.65 $\pm$ 0.95	-8.63 $\pm$ 1.48	1.1 $\pm$ 0.21	3.26 $\pm$ 0.61	5.36 $\pm$ 1.00	7.92 $\pm$ 1.49	Y = 2.76X - 0.12	0.97			
	D <sub>2%</sub> gEUD	-1.05 $\pm$ 0.23	-3.33 $\pm$ 0.74	-5.86 $\pm$ 1.25	-9.21 $\pm$ 1.85	0.99 $\pm$ 0.22	2.80 $\pm$ 0.60	4.45 $\pm$ 0.94	6.32 $\pm$ 1.31	Y = 2.58X - 0.54	0.95			
SP_PRV	D <sub>98%</sub>	-0.74 $\pm$ 0.13	-2.11 $\pm$ 0.39	-3.42 $\pm$ 0.63	-4.99 $\pm$ 0.90	0.78 $\pm$ 0.14	2.45 $\pm$ 0.47	4.28 $\pm$ 0.85	6.71 $\pm$ 1.38	Y = 1.94X + 0.33	0.96			
	D <sub>2%</sub> gEUD	-1.02 $\pm$ 0.48	-3.14 $\pm$ 1.50	-5.36 $\pm$ 2.40	-8.62 $\pm$ 3.56	1.01 $\pm$ 0.48	2.94 $\pm$ 1.35	4.76 $\pm$ 2.07	6.91 $\pm$ 2.81	Y = 2.53X - 0.24	0.85			
Parotid_L	D <sub>max</sub>	-1.52 $\pm$ 0.25	-4.46 $\pm$ 0.61	-7.32 $\pm$ 0.91	-10.67 $\pm$ 1.34	1.59 $\pm$ 0.35	4.83 $\pm$ 1.01	8.21 $\pm$ 1.58	12.5 $\pm$ 2.25	Y = 3.87X + 0.35	0.97			
	D <sub>2%</sub> gEUD	-1.56 $\pm$ 0.17	-4.57 $\pm$ 0.51	-7.46 $\pm$ 0.86	-10.93 $\pm$ 1.28	1.63 $\pm$ 0.18	4.92 $\pm$ 0.55	8.26 $\pm$ 0.93	12.51 $\pm$ 1.42	Y = 3.92X + 0.31	0.99			
Parotid_R	D <sub>max</sub>	-1.61 $\pm$ 0.36	-4.62 $\pm$ 1.06	-7.29 $\pm$ 1.69	-10.15 $\pm$ 2.24	1.78 $\pm$ 0.28	5.67 $\pm$ 0.75	9.88 $\pm$ 1.40	15.56 $\pm$ 2.28	Y = 4.29 + 1.02	0.96			
	D <sub>2%</sub> gEUD	-1.22 $\pm$ 0.21	-3.53 $\pm$ 0.66	-5.70 $\pm$ 1.08	-8.25 $\pm$ 1.62	1.29 $\pm$ 0.21	3.99 $\pm$ 0.63	6.87 $\pm$ 1.00	10.75 $\pm$ 1.34	Y = 3.16X + 0.47	0.97			
Parotid_R	D <sub>2%</sub> gEUD	-1.73 $\pm$ 0.26	-5.13 $\pm$ 0.78	-8.50 $\pm$ 1.29	-12.62 $\pm$ 1.91	1.77 $\pm$ 0.27	5.32 $\pm$ 0.82	8.90 $\pm$ 1.37	13.42 $\pm$ 2.07	Y = 4.34X + 0.16	0.98			
	D <sub>2%</sub> gEUD	-1.67 $\pm$ 0.28	-4.97 $\pm$ 0.82	-8.24 $\pm$ 1.32	-12.25 $\pm$ 1.93	1.72 $\pm$ 0.29	5.18 $\pm$ 0.89	8.66 $\pm$ 1.50	13.07 $\pm$ 2.28	Y = 4.22X + 0.17	0.97			

<sup>a</sup>%DE using the original plan as the baseline.

Abbreviations: BR\_PRV, planning organ at risk volume of brainstem; gEUD, generalized equivalent uniform dose; PGTVnd, planning target of involved lymph nodes; PGTVnx, planning target of gross tumor volume of the nasopharynx; PTV1, planning target of high-risk clinical target volume; PTV2, planning target of low-risk and preventive radiation volume; SP\_PRV, planning organ at risk volume of spinal cord.



**Figure 2.** The average gamma pass rate (%GP, mean  $\pm$  standard deviation [SD]) at different multileaf collimators (MLC) aperture errors of ArcCHECK (AC), portal dosimetry (PD), and logfile-based system (LF) at 3%/3mm (a), 3%/2mm (b), and 2%/2mm (c) criteria.

**Table 2.** The Gradient (slope) of Linear Fitting Between Gamma Pass Rate (%GP) and Multileaf Collimators (MLC) Closing/Opening Error for ArcCHECK (AC), Portal Dosimetry (PD) and Logfile-Based System (LF) at Different Dose Difference (DD)/Distance to Agreement (DTA) Criteria.

DD/DTA criteria	Slope at MLC aperture closing error (%/mm)			Slope at MLC aperture opening error (%/mm)		
	AC	PD	LF	AC	PD	LF
3%/3 mm	6.28	7.89	3.52	-6.58	-7.18	-4.06
3%/2 mm	9.25	12.29	6.64	-9.50	-11.37	-7.01
2%/2 mm	15.15	15.16	10.40	-16.94	-15.12	-10.58

targets (including all  $D_{98\%}$ ,  $D_{2\%}$ , and gEUD) in each MLC error plan was sorted at 1% intervals. For example, the %DE of 1% to 2% means the average of the absolute value of the percentage change in  $D_{98\%}$ ,  $D_{2\%}$ , and gEUD of all targets in a certain plan is  $\geq 1\%$  and  $< 2\%$ . The original plan without introducing error was regarded as a negative plan, and the plan with introducing error was regarded as a positive plan. The ROC curves, their corresponding AUC, and the cut-off values of the 3 DD/DTA criteria were calculated for different absolute MLC aperture errors or average absolute %DE of targets at 1% interval.

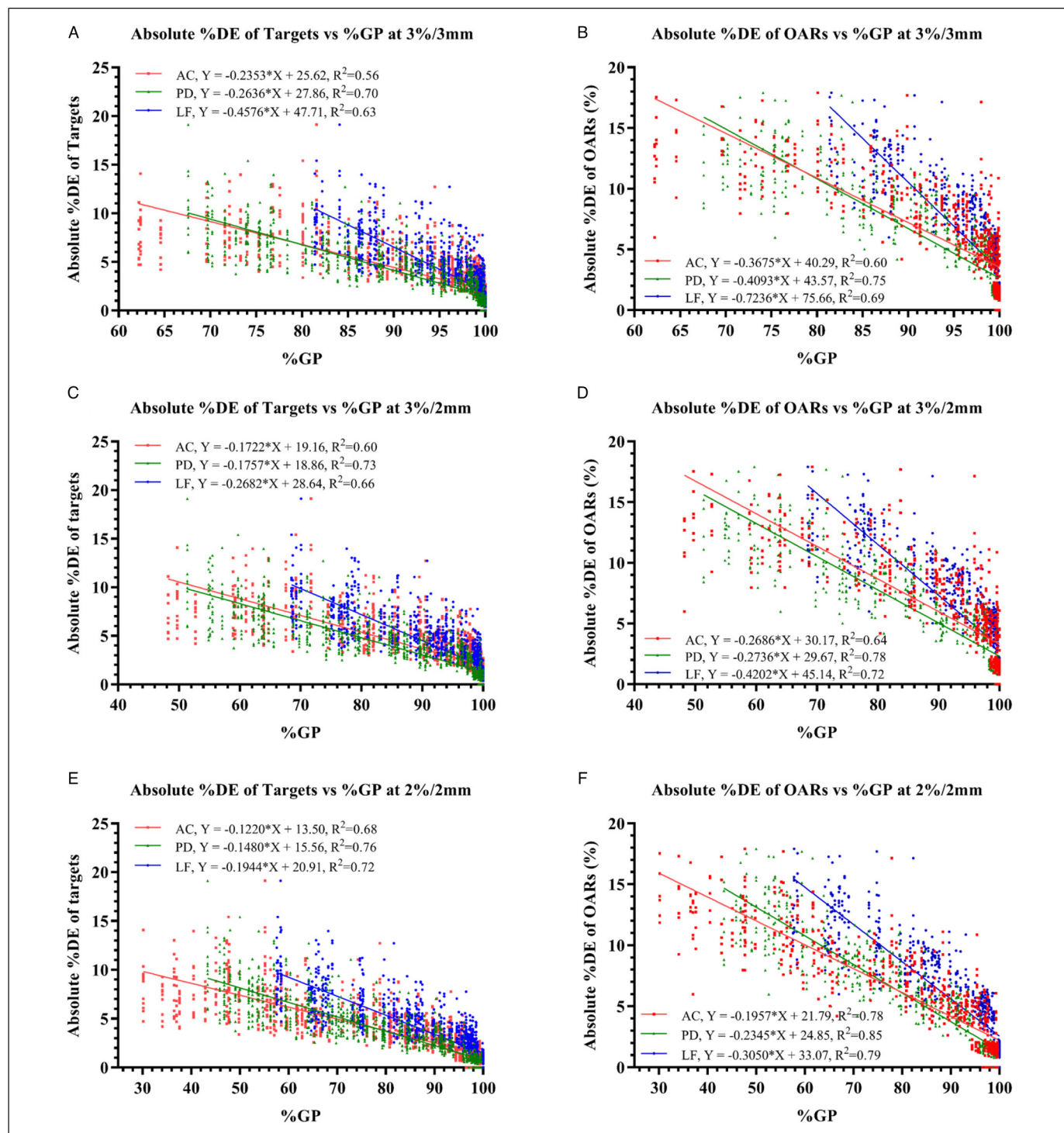
All data were analyzed using GraphPad Prism 8 (GraphPad Software Inc.).

## Results

The %DE between the plans with MLC aperture error and the original plan is shown in Table 1. The %DE of the dose metrics of OARs was larger than the %DE of the dose metrics of the targets of most cases. Gradient analysis shows that the changes of  $D_{98\%}$  of the target varied from 1.93% to 3.22% per millimeter of MLC aperture error. Changes of  $D_{2\%}$  and gEUD of target were 1.63%/mm to 2.47%/mm and 1.82%/mm to 2.76%/mm. Changes in gEUD of OARs were 3.16%/mm to 4.34%/mm. Changes in  $D_{max}$  of BR\_PRV and SP\_PRV were 3.87%/mm and 4.29%/mm.

The average %GP at different MLC aperture errors and DD/DTA criteria is shown in Figure 2. With the error increase, the %GP of the 3 systems has a significant difference. For the mean and standard deviation (SD) of %GP, LF had a larger mean and smaller SD in all the cases compared with the other 2 systems. The mean and SD of AC were larger than PD in most cases. The slope of linear fitting between %GP and MLC aperture error at different DD/DTA criteria is shown in Table 2. Compared with the other 2 systems, LF had the smallest absolute slope in any case. The absolute slope of PD was larger than AC, except for the case of opening error at the 2%/2mm criteria, where the slope of PD was  $-15.12\%/mm$  and the slope of AC was  $-16.94\%/mm$ .

The correlation between the %GP at different DD/DTA criteria and the absolute %DE including all dose metrics of all

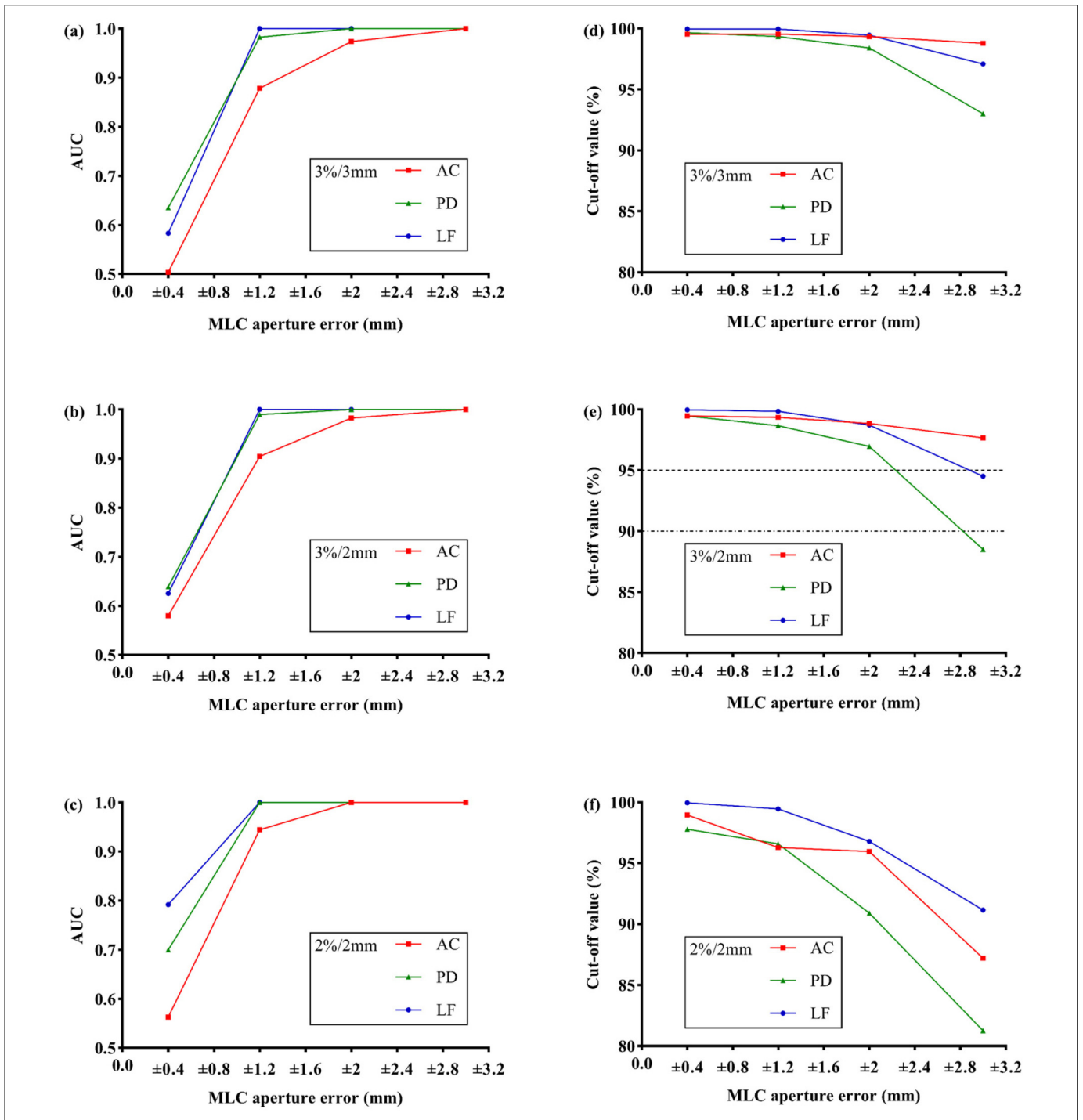


**Figure 3.** The correlation between the gamma pass rate (%GP) at different dose difference (DD)/distance to agreement (DTA) criteria ((A, B) 3%/3mm criteria, (C, D) 3%/2mm criteria, (E, F) 2%/2mm criteria) and the absolute value of percent dose error (%DE) including all dose metrics of all targets (A, C, E) or OARs (B, D, F).

targets or all OARs is shown in Figure 3. In all the cases, the range of  $R^2$  was 0.56 to 0.85. PD has the largest  $R^2$ , followed by LF, and AC the smallest. The  $R^2$  of all cases increased gradually as the DD/DTA criterion became stricter. The data was scattered, %GP does not fully represent the level of %DE,

and there were cases where both %DE and %GP were large, which was obvious for the case of OARs.

Figure 4 shows the AUC and optimal cut-off value calculated from the ROC curve of different systems as a function of different absolute MLC aperture errors at different DD/

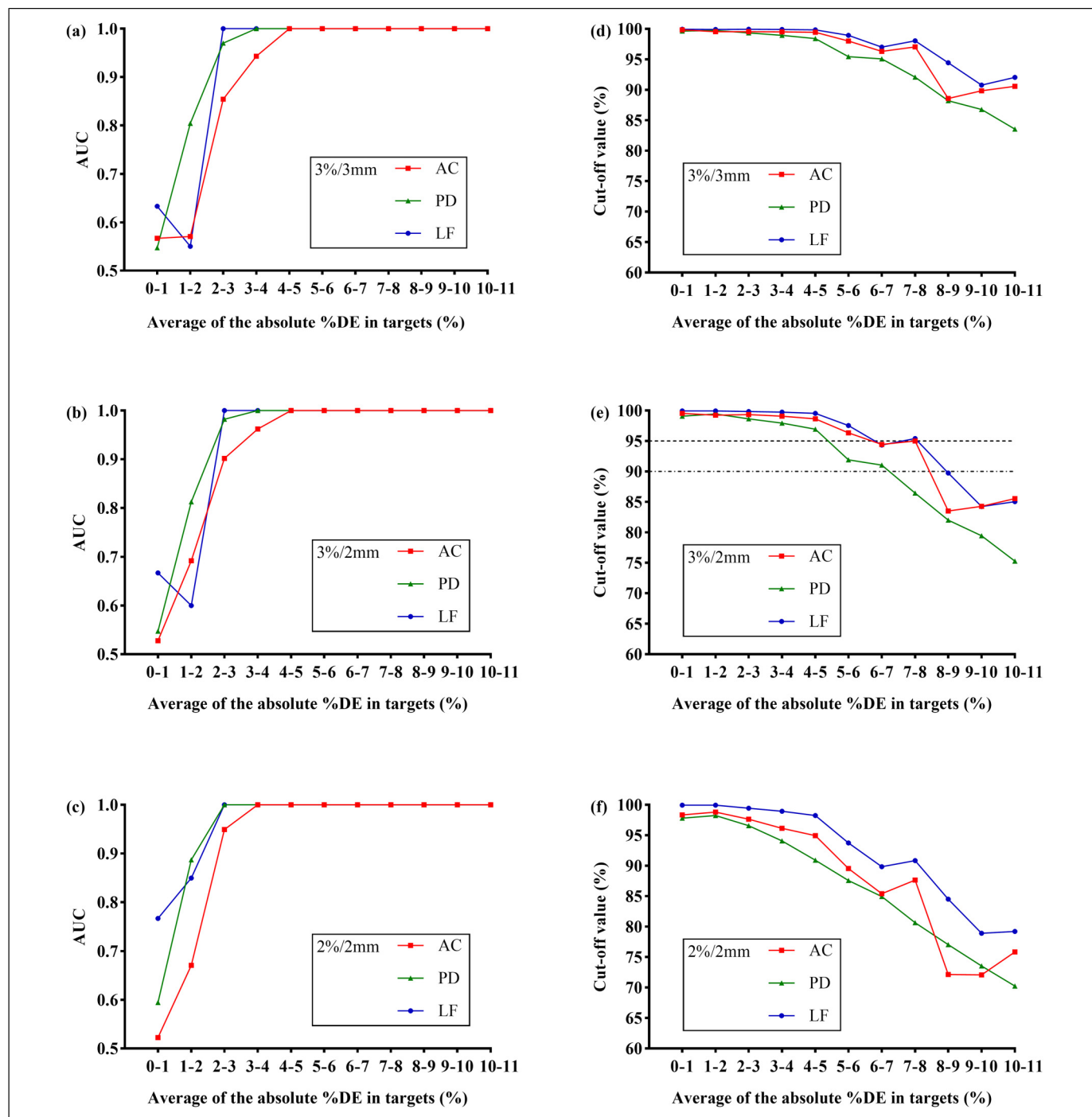


**Figure 4.** The area under the receiver operating characteristic (ROC) curve (AUC) and optimal cut-off value of ArcCHECK (AC), portal dosimetry (PD), and logfile-based system (LF) as a function of multileaf collimators (MLC) aperture error at 3%/3mm (A, D), 3%/2mm (B, E), and 2%/2mm (C, F) criteria.

DTA criteria. Before the AUC value of all systems reaches 1, the AUC of AC was the smallest among the 3 systems. The AUC of most cases increased gradually as the DD/DTA criterion became stricter. When the error was  $\pm 0.4$  mm, the AUC value of LF and PD at 2%/2mm criteria were 0.79 and 0.7, and the AUC values of the other cases were all  $< 0.7$ , of which AC

was  $< 0.6$ . When the error was  $\pm 1.2$  mm, the AUC value of AC at 3%/3mm criteria was 0.88, and the AUC values of the other cases were all larger than 0.9. The AUC value increases as the error increases, and when it is equal to 1, it means that the system can perfectly identify the error. For the 3%/3mm criteria (Figure 4 (a,d)), the error that AC, PD,





**Figure 5.** The area under the receiver operating characteristic (ROC) curve (AUC) and optimal cut-off value of ArcCHECK (AC), portal dosimetry (PD) and logfile-based system (LF) as a function of the average absolute value of percent dose error (%DE) in targets at 3%/3mm (A, D), 3%/2mm (B, E), and 2%/2mm (C, F) criteria. The ROC curve was obtained by considering the original plans as negative and the plans within different %DE ranges as positive, respectively. For example, the %DE range from 1% to 2% means the average of the absolute value of the percentage change in  $D_{98\%}$ ,  $D_{2\%}$ , and generalized equivalent uniform dose (gEUD) of the target is  $\geq 1\%$  and  $< 2\%$ .

and LF can perfectly identify were  $\pm 3\text{mm}$ ,  $\pm 2\text{mm}$ , and  $\pm 1.2\text{mm}$ , respectively, and the corresponding optimal cut-off values were 98.8%, 98.4%, and 99.95%, respectively. For the 3%/2mm criteria (Figure 4 (b,e)), the error that AC, PD, and LF can perfectly identify were the same as the 3%/3mm criteria, and the

corresponding optimal cut-off values were 97.65%, 96.95%, and 99.85%, respectively. For the 2%/2mm criteria (Figure 4 (c,f)), AC can perfectly identify the error of  $\pm 2\text{mm}$ , and PD and LF can identify the error of  $\pm 1.2\text{mm}$ . The corresponding optimal cut-off values were 95.95%, 96.6%, and 99.45%, respectively.

Figure 5 shows AUC and the optimal cut-off value calculated from the ROC curve of different systems as a function of the average absolute %DE with a 1% interval in targets at different DD/DTA criteria. Before the AUC value of all systems reaches 1, the AUC of AC was the smallest among the 3 systems except for some cases where %DE  $\leq$  2%. For the 3%/3mm criteria (Figure 5 (a,d)), when the %DE were 0% to 1% and 1% to 2%, the AUC of all systems were all-around 0.6 but for the AUC of PD at 1% to 2% was 0.8. When the %DE was 2% to 3%, all systems had an AUC larger than 0.85. The %DE that AC, PD, and LF can perfectly identify were 4% to 5%, 3% to 4%, and 2% to 3%, respectively, and the corresponding optimal cut-off values were 99.45%, 98.95%, and 99.95%, respectively. For the 3%/2mm criteria (Figure 5 (b,e)), the AUC of all systems was slightly larger than the 3%/3mm criteria except for the AUC of AC when %DE is 0% to 1%. However, the %DE that AC, PD, and LF can perfectly identify were the same as the 3%/3mm criteria, and the corresponding optimal cut-off values were 98.65%, 97.95%, and 99.85%, respectively. For the 2%/2mm criteria (Figure 5 (c,f)), PD and LF showed large improvement in the ability to identify dose error of 1% to 2%, where the AUC were 0.89 and 0.85. AC can perfectly identify the %DE of 3% to 4%, and PD and LF can identify the %DE of 2% to 3%. The corresponding optimal cut-off values were 96.15%, 96.6%, and 99.45%, respectively.

## Discussion

In this study, the impact of MLC aperture errors on dose distribution of head and neck VMAT plans was analyzed, and the sensitivity of 3 PSQA systems to these errors was compared. The TG-218 suggested that dose verification should use the method that best identifies errors. True composite method was recommended for AC, but per beam analysis was recommended for EPID because composite field may mask errors. However, the simulated error in this study was MLC opening and closing error, and the composite field will not mask this type of error but may be more conducive to the identification of this type of error, so the composite field was used for comparison in all systems. The results show that the MLC aperture error has a large effect on the dose distribution of the VMAT plan. The average %DE of most structures was  $\geq$  3% when the MLC aperture error was  $\geq$  1.2mm. There was a large difference in %GP among the 3 systems. PD was more sensitive to MLC aperture errors, followed by AC, and LF was the least sensitive. LF had a larger mean but smaller SD and slope compared with the other 2 systems at different DD/DTA criteria. Compared with AC, PD had a smaller mean and SD but a larger slope in most cases. The high %GP of logfile-based dose verification was reported in the study of Stanislav et al,<sup>25</sup> which compared the %GP between a LF using python script API Pylinac and the PD system by changing the dosimetric leaf gap (DLG). They found that the results of the LF were "too good" compared with PD while using the DD/DTA criteria of 3%/1mm, 2%/2mm, and 1%/1mm, respectively. To make the

results of these 2 systems more comparable, they found that the dose difference criterion used for the LF should be 8 to 10 times smaller than the criterion used with PD. The LF was not affected by factors such as detector and setup error, which explains the better results of the LF compared to the other systems. Furthermore, QA based on a logfile can analyze each delivery parameter and find small delivery errors that cannot be detected by the phantom-based measurement.<sup>26,27</sup> Despite several advantages, replacing measurement-based QA with logfile-based QA remains controversial among physicists.<sup>28</sup> Logfile is not considered an independent measurement tool since the recorded data is measured by LINAC and not by a third party. The logfile cannot record some errors caused by machine calibration, hardware malfunction, etc. Agnew et al<sup>14</sup> reported that the MLC position recorded on the logfile was inconsistent with the measurement results of EPID due to a loose T-nut. Strict machine QA for LINAC should be performed before logfiles are used for PSQA.

The data was scattered in correlation analysis between the %GP at different DD/DTA criteria and the absolute %DE including all dose metrics of all targets or all OARs. A %GP may correspond to many different %DE, and a %DE may correspond to many different %GP. This is because the impact of the error on different structures and dose metrics is not the same. The same error may have a larger impact on some structures and dose metrics and a smaller impact on others. %GP does not fully represent the level of %DE, and there were cases where both %DE and %GP were large. The effectiveness of the gamma evaluation method for pre-treatment PSQA has been questioned by researchers, with some studies suggesting that %GP does not reflect clinical outcome.<sup>18,29,30</sup> The gamma index is more about the accuracy of machine delivery rather than the dose distribution of the patient in that it does not indicate the location and distribution of the non-passing points in the patient. For example, if the %GP was 95% with 3%/2mm criteria, usually it was considered the plan could be used for clinical treatment. However, if it were known that the failed 5% are concentrated on the spinal cord, it could lead to a serious impact on its maximum dose. At this time, it is important to consider whether the plan can be used for treatment. McKenzie et al<sup>18</sup> investigated the ability of several PSQA systems to identify acceptable and unacceptable intensity-modulated radiation therapy (IMRT) plans. They found that most systems performed poorly at identifying unacceptable plans. However, most of these studies are based on looser %GP limits. In fact, whether a plan is accepted or not has much to do with the acceptable error size and the corresponding %GP limit.

Carlone et al<sup>16</sup> investigated the use of the ROC curve to get tolerance limits for gamma evaluation using MapCHECK by introducing random MLC errors to the prostate IMRT plan. They found that the gamma criterion achieved nearly 100% sensitivity/specificity for random errors larger than 3mm, and the optimal cut-off values for different error detection range from 92% to 99% for 3%/3 mm criteria, whereas for 2%/2 mm, the range was from 77% to 94%. In this study, the AUCs and cut-off values were obtained from ROC curves of 3 PSQA

systems at different DD/DTA criteria for different absolute MLC aperture errors or average absolute %DE of targets. Although the %GP of LF was generally high, it had good stability (small SD value) at different errors; therefore, it was easier to find a suitable limit to distinguish plans with and without errors. Compared with the other 2 systems, the AUC of LF can reach 1 faster in most cases, indicating that the error of perfect identification is smaller. Except for LF at 2%/2mm criteria, all systems had poor recognition ability for  $\pm 0.4$ mm MLC aperture error or 0% to 1% dose error, with AUC less than 0.7. AC has the worst ability to distinguish plans with and without errors in most cases. For MLC aperture error, the error of AC to achieve perfect discriminative ability appears to be large. This may be related not only to the characteristics of the AC itself but also to the plan. First, in terms of the characteristics of the AC itself, it is a lot harder to perfectly predict the doses to be measured in every detector, considering they are irradiated with both entrance and exit doses and a lot more scattered and oblique doses than an EPID where all is a perfectly perpendicular entrance dose. In addition, EPID has a much higher spatial resolution than AC. Therefore, there may be larger uncertainty when using AC to verify plans with large errors. The %GP vary over a wide range for different plans with large errors making it more difficult to find an appropriate limit to distinguish plans with and without errors. Second, the error of the same size has different effects on different plans. In this study, we only simulated the MLC aperture error without simulating the error of the jaws. The same open error may have a smaller impact on plans with more leaves close to the edge of the jaws. Furthermore, the impact of MLC errors on plans with a large average segment width may be smaller. Moon et al<sup>31</sup> investigated the correlation between average segment width and %GP as a function of MLC position error. They confirmed a strong correlation between the %GP descent gradient due to MLC position error and average segment width. MLC position errors have less effect on plans with larger average segment width, so we can reduce the effect of MLC position errors on dose delivery uncertainty by limiting the plan's minimum segment width. Therefore, for all verification systems, it is possible that the pass rate of individual plans is still high when the MLC error is large, which may lead to the failure of finding a perfect limit to distinguish plans with and without errors. For %DE, there are many dose metrics of many different structures that can be used to evaluate the quality of a plan, and it is difficult to decide which one or several metrics of which structures are used as the evaluation objects. In this study, ROC analysis was performed using the average absolute %DE of  $D_{0.8}$ ,  $D_2$ , and  $D_{\text{mean}}$  in all targets of a certain plan, regardless of the dose metrics of OARs. The effect of MLC aperture errors on the dose metrics of OAR was larger than on the dose metrics of the target. It is more difficult to determine the magnitude of acceptable %DE of OAR than the target since even a large %DE may be acceptable when the dose metrics of OAR are far below the clinical limit.

It can be seen from Figures 4(e) and 5(e) that some large errors may be omitted if the universal tolerance limits (95%)

are adopted in the case of this study. When the cut-off value is set on the universal tolerance limit, the MLC aperture error that can be well identified by AC, PD, and LF is larger than 3mm, approximately 2mm, and approximately 3mm, respectively. And the %DE that can be well identified by PD is approximately 4% to 5%, while AC and LF are approximately 6% to 8%. This suggests that the universal tolerance limits are not necessarily applicable to all situations in different institutions. In some cases, it may not be sensitive enough to detect clinically significant errors, resulting in potential risks to the PSQA process. We usually focus more on the %GP of PSQA systems than on their ability to identify errors. In fact, the %GP is affected by many factors, such as different TPS, LINACS, plan types, and even different quality control personnel may bring about alternative results. A high %GP does not necessarily mean that the plan is error-free, and a low %GP does not mean that the plan contains errors. Therefore, it may not be appropriate to use the same gamma analysis criteria and limits for all PSQA systems. What is more important than a high or low %GP is finding an appropriate limit to distinguish the plans with errors. The %GP of PD and LF were significantly different, but they could all obtain good error identification ability using an appropriate limit value. Each institution should formulate its own customized %GP limit based on its PSQA process. ROC curve is one of the common research methods used to reflect the ability to discriminate between negative and positive aspects of a certain index in clinical research. However, the premise of its application is that a group of known negative and positive plans is required. In this study, only the MLC opening and closing errors, which have large clinical impact, were simulated in head and neck VMAT plans. More error types such as machine output, gantry speed, and other treatment sites or modalities such as SBRT/SRS need to be investigated in future studies. In addition, the method based on statistical process control is also considered an effective way to find a suitable limit to detect errors in the PSQA process.<sup>15,32,33</sup> The advantages and disadvantages of this process can be determined by comparing it with the ROC method in future studies.

## Conclusions

Gamma evaluation can be greatly affected by QA systems and user-defined parameters. The sensitivity of 3 PSQA systems to MLC aperture errors of head and neck VMAT plan was investigated in this study. The gradient analysis showed that LF had the lowest sensitivity to MLC aperture errors. The sensitivity of PD was higher than AC, except for the case of open error at 2%/2mm criteria. However, the ROC analysis demonstrated that the AC had the lowest ability to identify plans with errors in most cases. The error that LF can fully identify is usually smaller than the other 2 systems. The error identification ability usually increased gradually as the DD/DTA criterion became stricter. %GP does not fully represent the level of %DE. We should be more concerned with finding an appropriate limit to identify plans with errors rather than just pursuing a high %GP thus

institutions should formulate their own customized limits based on their PSQA process through ROC or other methods.


### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Key Research and Development Program of China (grant No. 2016YFC0105103).

### ORCID iD

Jinyou Hu  <https://orcid.org/0000-0001-5118-4725>

### References

- Moran JM, Dempsey M, Eisbruch A, et al. Safety considerations for IMRT: executive summary. *Pract Radiat Oncol.* 2011;1(3):190-195. <https://doi.org/10.1016/j.prro.2011.04.008>
- Palma D, Vollans E, James K, et al. Volumetric modulated arc therapy for delivery of prostate radiotherapy: comparison with intensity-modulated radiotherapy and three-dimensional conformal radiotherapy. *Int J Radiat Oncol Biol Phys.* 2008;72(4):996-1001. <https://doi.org/10.1016/j.ijrobp.2008.02.047>
- Verbakel WF, Cuijpers JP, Hoffmans D, Bieker M, Slotman BJ, Senan S. Volumetric intensity-modulated arc therapy vs. conventional IMRT in head-and-neck cancer: a comparative planning and dosimetric study. *Int J Radiat Oncol Biol Phys.* 2009;74(1):252-259. doi:10.1016/j.ijrobp.2008.12.033
- Wolff D, Stieler F, Welzel G, et al. Volumetric modulated arc therapy (VMAT) vs. serial tomotherapy, step-and-shoot IMRT and 3D-conformal RT for treatment of prostate cancer. *Radiother Oncol.* 2009;93(2):226-233. doi:10.1016/j.radonc.2009.08.011
- Davidson MT, Blake SJ, Batchelar DL, Cheung P, Mah K. Assessing the role of volumetric modulated arc therapy (VMAT) relative to IMRT and helical tomotherapy in the management of localized, locally advanced, and post-operative prostate cancer. *Int J Radiat Oncol Biol Phys.* 2011;80(5):1550-1558. doi:10.1016/j.ijrobp.2010.10.024
- Oliver M, Gagne I, Bush K, Zavgorodni S, Ansbacher W, Beckham W. Clinical significance of multi-leaf collimator positional errors for volumetric modulated arc therapy. *Radiother Oncol.* 2010;97(3):554-560. <https://doi.org/10.1016/j.radonc.2010.06.013>
- Bai S, Li G, Wang M, Jiang Q, Zhang Y, Wei Y. Effect of MLC leaf position, collimator rotation angle, and gantry rotation angle errors on intensity-modulated radiotherapy plans for nasopharyngeal carcinoma. *Med Dosim.* 2013;38(2):143-147. doi:10.1016/j.meddos.2012.10.002
- Blake SJ, Arumugam S, Holloway L, et al. Investigating the impact of treatment delivery uncertainties on treatment effectiveness for lung SABR. *Australas Phys Eng Sci Med.* 2017;40(4):823-829. doi:10.1007/s13246-017-0591-x
- Pogson EM, Aruguman S, Hansen CR, et al. Multi-institutional comparison of simulated treatment delivery errors in ssIMRT, manually planned VMAT and autoplan-VMAT plans for nasopharyngeal radiotherapy. *Phys Med.* 2017;42:55-66. doi:10.1016/j.ejmp.2017.08.008
- Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys.* 1998;25(5):656-661. <https://doi.org/10.1118/1.598248>
- Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. *Med Phys.* 2003;30(9):2455-2464. <https://doi.org/10.1118/1.1598711>
- Sun B, Rangaraj D, Palaniswaamy G, et al. Initial experience with TrueBeam trajectory log files for radiation therapy delivery verification. *Pract Radiat Oncol.* 2013;3(4):e199-e208. <https://doi.org/10.1016/j.prro.2012.11.013>
- Chow V, Kan M, Chan A. Patient-specific quality assurance using machine log files analysis for stereotactic body radiation therapy (SBRT). *J Appl Clin Med Phys.* 2020;21:179-187. doi:10.1002/acm2.13053
- Agnew A, Agnew C, Grattan M, Hounsell A, McGarry C. Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries. *Phys Med Biol.* 2014;59(9):49-63. doi:10.1088/0031-9155/59/9/N49
- Miften M, Olch A, Mihailidis D, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: recommendations of AAPM task group No. 218. *Med Phys.* 2018;45(4):e53-e83. <https://doi.org/10.1002/mp.12810>
- Carlone M, Cruje C, Rangel A, McCabe R, Nielsen M, MacPherson M. ROC Analysis in patient specific quality assurance. *Med Phys.* 2013;40(4):042103. <https://doi.org/10.1118/1.4795757>
- Pogson EM, Arumugam S, Hansen CR, et al. Comparison of multi-institutional pre-treatment verification for VMAT of nasopharynx with delivery errors. *Phys Med.* 2018;53:25-31. doi:10.1016/j.ejmp.2018.07.007
- McKenzie EM, Balter PA, Stingo FC, Jones J, Followill DS, Kry SF. Toward optimizing patient-specific IMRT QA techniques in the accurate detection of dosimetrically acceptable and unacceptable patient plans. *Med Phys.* 2014;41(12):121702. doi:10.1118/1.4899177
- Bojchko C, Ford EC. Quantifying the performance of in vivo portal dosimetry in detecting four types of treatment parameter variations. *Med Phys.* 2015;42(12):6912-6918. doi:10.1118/1.4935093
- Niemierko A. A generalized concept of equivalent uniform dose (abstr.). *Med Phys.* 1999;26:1101.
- Allen Li X, Alber M, Deasy JO, et al. The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM. *Med Phys.* 2012;39:1386-1409. <https://doi.org/10.1118/1.3685447>
- Kerns J. Pylinac API Documentation version 2.5.1. [https://pylinac.readthedocs.io/en/v2.5.1/log\\_analyzer.html](https://pylinac.readthedocs.io/en/v2.5.1/log_analyzer.html).
- Peet SC. Flashgamma. <https://github.com/samuelpet/flashgamma>.
- Peet SC, Yu L, Maxwell S, Crowe SB, Trapp JV, Kaim T. Exploring the gamma surface: a new method for visualising modulated radiotherapy quality assurance results. *Phys Med.* 2020;78:166-172. <https://doi.org/10.1016/j.ejmp.2020.09.021>
- Bozhikov S, Mitarova K, Vassileva F, Tacheva B, Paarvanova B, Karabaliev M. Determination of the sensitivity of trajectory log

- files gamma analysis - A method based on python scripts. *AIP Conf Proc*. 2019;2186:110004. doi: 10.1063/1.5138023
26. Rangaraj D, Zhu M, Yang D, et al. Catching errors with patient-specific pretreatment machine log file analysis. *Pract Radiat Oncol*. 2013;3(2):80-90. doi:10.1016/j.prro.2012.05.002
27. Chow VUY, Kan MWK, Chan ATC. Patient-specific quality assurance using machine log files analysis for stereotactic body radiation therapy (SBRT). *J Appl Clin Med Phys*. 2020;21(11):179-187. doi:10.1002/acm2.13053
28. Childress N, Chen Q, Rong Y. Parallel/opposed: IMRT QA using treatment log files is superior to conventional measurement-based method. *J Appl Clin Med Phys*. 2015;16:5385. doi:10.1120/jacmp.v16i1.5385
29. Nelms BE, Zhen H, Tomé WA. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. *Med Phys*. 2011;38(2):1037-1044. <https://doi.org/10.1118/1.3544657>
30. Stasi M, Bresciani S, Miranti A, Maggio A, Sapino V, Gabriele P. Pretreatment patient-specific IMRT quality assurance: a correlation study between gamma index and patient clinical dose volume histogram. *Med Phys*. 2012;39:7626-7634. doi:10.1118/1.4767763
31. Moon YM, Bae SI, Han MJ, et al. Correlation between average segment width and gamma passing rate as a function of MLC position error in volumetric modulated arc therapy. *Technol Cancer Res Treat*. 2021;20:1-10. doi:10.1177/15330338211059937
32. Pawlicki T, Whitaker M, Boyer AL. Statistical process control for radiotherapy quality assurance. *Med Phys*. 2005;32(9):2777-2786. <https://doi.org/10.1118/1.2001209>
33. Breen SL, Moseley DJ, Zhang B, Sharpe MB. Statistical process control for IMRT dosimetric verification. *Med Phys*. 2008;35(10):4417-4425. <https://doi.org/10.1118/1.2975144>