



Assessment of Statistical Process Control Based DVH Action Levels for Systematic Multi-Leaf Collimator Errors in Cervical Cancer RapidArc Plans

OPEN ACCESS

Hanyin Zhang[†], Wenli Lu[†], Haixia Cui, Ying Li and Xin Yi^{*}

Department of Oncology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China

Edited by:

Jason W. Sohn,
Allegheny Health Network,
United States

Reviewed by:

Yiran Zheng,
University Hospitals Cleveland Medical
Center, United States
Prabakar Sukumar,
Oxford University Hospitals,
United Kingdom
Mahmoud Ahmed,
Vanderbilt University Medical Center,
United States

*Correspondence:

Xin Yi
noodles@hospital.cqmu.edu.cn

[†]These authors have contributed
equally to this work and share
first authorship

Specialty section:

This article was submitted to
Radiation Oncology,
a section of the journal
Frontiers in Oncology

Received: 26 January 2022

Accepted: 19 April 2022

Published: 18 May 2022

Citation:

Zhang H, Lu W, Cui H, Li Y and Yi X
(2022) Assessment of Statistical
Process Control Based DVH
Action Levels for Systematic Multi-
Leaf Collimator Errors in Cervical
Cancer RapidArc Plans.
Front. Oncol. 12:862635.
doi: 10.3389/fonc.2022.862635

Background: In the patient-specific quality assurance (QA), DVH is a critical clinically relevant parameter that is finally used to determine the safety and effectiveness of radiotherapy. However, a consensus on DVH-based action levels has not been reached yet. The aim of this study is to explore reasonable DVH-based action levels and optimal DVH metrics in detecting systematic MLC errors for cervical cancer RapidArc plans.

Methods: In this study, a total of 148 cervical cancer RapidArc plans were selected and measured with COMPASS 3D dosimetry system. Firstly, the patient-specific QA results of 110 RapidArc plans were retrospectively reviewed. Then, DVH-based action limits (AL) and tolerance limits (TL) were obtained by statistical process control. Secondly, systematic MLC errors were introduced in 20 RapidArc plans, generating 380 modified plans. Then, the dose difference (%DE) in DVH metrics between modified plans and original plans was extracted from measurement results. After that, the linear regression model was used to investigate the detection limits of DVH-based action levels between % DE and systematic MLC errors. Finally, a total of 180 test plans (including 162 error-introduced plans and 18 original plans) were prepared for validation. The error detection rate of DVH-based action levels was compared in different DVH metrics of 180 test plans.

Results: A linear correlation was found between systematic MLC errors and %DE in all DVH metrics. Based on linear regression model, the systematic MLC errors between -0.94 mm and 0.88 mm could be caught by the TL of PTV₉₅ ([-1.54%, 1.51%]), and the systematic MLC errors between -1.00 mm and 0.80 mm could also be caught by the TL of PTV_{mean} ([-2.06%, 0.38%]). In the validation, for original plans, PTV₉₅ showed the minimum error detection rate of 5.56%. For error-introduced plans with systematic MLC errors more than 1mm, PTV_{mean} showed the maximum error detection rate of 88.89%, and then was followed by PTV₉₅ (86.67%). All the TL of DVH metrics showed a poor error detection rate in identifying error-induced plans with systematic MLC errors less than 1mm.

Conclusion: In 3D quality assurance of cervical cancer RapidArc plans, process-based tolerance limits showed greater advantages in distinguishing plans introduced with systematic MLC errors more than 1mm, and reasonable DVH-based action levels can be acquired through statistical process control. During DVH-based verification, main focus should be on the DVH metrics of target volume. OARs in low-dose regions were found to have a relatively higher dose sensitivity to smaller systematic MLC errors, but may be accompanied with higher false error detection rate.

Keywords: cervical cancer RapidArc plan, 3D quality assurance, DVH-based action levels, statistical process control, systematic MLC errors

INTRODUCTION

As a rapidly developed and increasingly used technique, volumetric modulated arc therapy (VMAT) is capable of delivering a high conformal dose distribution in a short delivery time (1, 2). As Varian's commercial implementation of the VMAT algorithm proposed by Otto in 2008 (3), RapidArc incorporates concomitant continuous gantry rotation, dynamic movement of multileaf collimators (MLC) and variable dose rate (4). Due to rotational delivery feature and increased complexity in planning and delivery, RapidArc poses a great challenge to treatment planning system (TPS) and linac performance. In order to ensure the safety of patient, patient-specific quality assurance (QA) must be implemented before the delivery of RapidArc plan. Action level is often used to determine whether the delivered plan is appropriate for the treatment of a patient. If the result of a patient-specific QA exceeds a predetermined action level, the plan will be delayed until the source of error is identified or the treatment is re-planned (5).

Nowadays, gamma analysis is the most widely used method for patient-specific QA. However, the results of gamma analysis have not been found to correlate well with clinically relevant metrics (such as the estimated deviations in dose volume histograms) (6). Under this circumstance, some researchers have incorporated DVH information into patient-specific QA results. In 2011, Nemls et al. (7) had proposed that "false negative" and "false positive" in conventional QA results could be revealed by DVH metrics. However, they failed to clarify how to set up action levels of DVH metrics reasonably and scientifically. In their following research, they have suggested that the difference of DVH metrics exceeding 5% represents a clinical implication and further studies about DVH-based action levels are required (8). Subsequently, different DVH-based action levels ranging from 2% to 5% have been proposed in a significant number of publications (4, 5, 9–11). Among them, 3% or 5% has been the most commonly used DVH-based action levels. In AAPM TG-119 report (12), the action levels of 4.5% and 4.7% have been recommended for point dose measurements in target and low-dose regions, respectively. In AAPM TG-218 report (13), it has been emphasized that universal action levels do not fit every institution, every case or every structure. In addition, some researchers (9, 14) have attempted to adopt specific DVH-based action levels for different structures, such as 3% for target volume

and 5% for OARs. Ruurd Visser et al. (15) have also set DVH-based action levels of 2%–4% to evaluate nine structure types in head and neck IMRT treatment plans. They have found that the proposed DVH-based action levels are too strict for some structures, such as target volume or OARs near the target volume. In our previous study (16), the most commonly used DVH-based action level (3% or 5%) was found to be unsuitable for all structures. Because QA results may be affected by QA equipment, delivery system, tumor sites, plan's complexity and so on. Given the above, it is a challenge to set a reasonable and scientific DVH-based action level according to actual situation.

Actually, AAPM TG-218 report has also focused on the issue that general action levels do not fit all situations, and recommended that locally defined action levels could be achieved by setting process-based tolerance and action limits. The locally defined action levels should ultimately be specific to local equipment, processes and case types as well as the experience of local physicist (17). Process-based tolerance and action limits are derived from statistical process control (SPC). SPC is a powerful analytical decision-making tool for monitoring production processes, achieving stability, and reducing variability (18). It mainly consists of two steps (19). The first step is to collect empirical data to determine the control limits, and the second step is to monitor the process by using a control chart to detect whether the data exceed the control limits. As an important process to ensure the safety of patients and fidelity of treatment, SPC could also be used to monitor the results of patient-specific QA. Actually, the application of SPC for analyzing QA process in radiotherapy has been growing over recent years and achieved good results (18–23). However, SPC is primarily utilized for obtaining locally defined action levels about gamma passing rates. No studies to date have rigorously validated the process-based DVH action levels for RapidArc plans. In addition, a variety of structures or DVH metrics are included in radiotherapy plan, so the process-based DVH action levels can also be calculated in different structures or DVH metrics. Therefore, process-based tolerance and action limits should be promising and viable options for obtaining DVH-based action levels.

In current radiotherapy practice, DVH is a critical clinically relevant parameter that is finally used to determine the safety and effectiveness of radiotherapy, and it has indeed improved the correlation between patient-specific QA results and clinically

relevant metrics, but a consensus on DVH-based action levels has not been reached yet. In this study, different magnitude of systematic MLC errors were introduced in cervical cancer RapidArc plans, and a comprehensive and systematic evaluation was performed between process-based DVH action levels and commonly used DVH-based action levels. By doing so, we aimed to explore reasonable DVH-based action levels and optimal DVH metrics in detecting systematic MLC errors.

MATERIALS AND METHODS

Overview

In this study, we designed a method to evaluate the detectability of different DVH-based action levels in systematic MLC errors. The method was summarized as below and schematized in **Figure 1**. It was mainly composed of three steps. Firstly, 110 RapidArc plans were selected and their patient-specific QA results were reviewed. DVH-based action levels were acquired by statistical process control. Meanwhile, other commonly used DVH-based action levels were also collected for comparison. Secondly, 20 other RapidArc plans were selected, in which different magnitude of systematic MLC errors were introduced. QA measurement was performed for all original and modified plans by COMPASS 3D dosimetry system. Then, the dose errors (%DE) of DVH information in specific structures were extracted from the measurement-based QA results of all original plans and modified plans. After that, the %DE of DVH metrics and the magnitude of systematic MLC errors were analyzed by linear regression analysis. The detection limits in DVH-based action levels mentioned in first step were investigated and compared by linear regression model to find out reasonable action levels for detecting systematic MLC errors. Thirdly, a total of 180 plans, including 162 error-introduced plans and 18 original plans, were prepared for validation. The error detection rate of DVH action

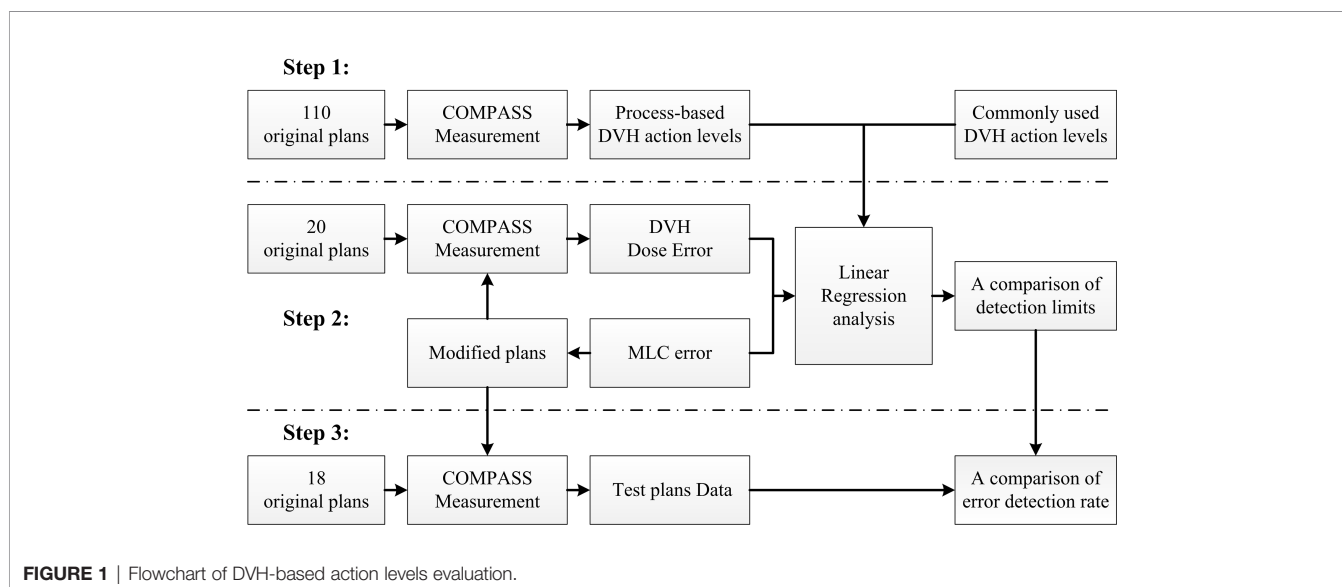
levels was compared in different DVH metrics in order to explore the optimal DVH metrics to detect systematic MLC errors.

Treatment Planning and Delivery

From 2017 to 2020, a total of 148 cervical cancer RapidArc plans were selected for this retrospective study. The aim of these enrolled plans was to deliver dose equal to 50Gy (in 25 fractions) to 95% of the planning target volume (PTV), while simultaneously meeting the plan acceptance for critical structures (bladder, rectum and femoral heads). All plans contained two full arcs: CCW with a start angle of 178° and collimator angle of 350°, and CW with a start angle of 182° and collimator angle of 10°. Each plan was calculated on a 2.5 mm isotropic dose grid with anisotropic analytical algorithm (AAA) through Eclipse v13.5 (Varian Medical Systems, Palo Alto, CA, USA). These plans were delivered by a 6 MV linear accelerator (Unique, Varian Medical Systems, Palo Alto, CA, USA) equipped with a millennium 120 multileaf collimators. All plans were delivered to COMPASS 3D dosimetry system for dose verification before patient treatment. In dose verification, the enrolled plans must meet the following criteria: 1. The gamma passing rates (GPRs) (3%/2mm, 10% dose threshold) should be greater than 95%. 2. The mean gamma index (GI) (3%/2mm, 50% dose threshold) should be smaller than 0.5.

COMPASS 3D Dosimetry System

The COMPASS 3D dosimetry system (IBA Dosimetry, Germany) was used to generate independent data for 3D dose verification. This system is composed of a MatriXX 2D array and the analysis software COMPASS V4.1. The MatriXX 2D array consists of 1020 vented pixel parallel plate ionization chambers with a spatial resolution of 7.62mm (center-to-center distance of chambers). In dose verification, dose distribution in patient CT could be determined by either a dose calculation (model-based) or a dose reconstruction (measurement-based).



In this study, QA data were focused on COMPASS measurement-based QA results. In the measurement-based QA, 3D dose distribution was reconstructed based on fluence measured by MaritXX 2D array and dose calculation was performed by collapsed cone convolution (CCC) algorithm on patient's CT. In this situation, data transfer and linac behavior were taken into account during comparing the TPS calculated dose with the COMPASS reconstructed dose. The QA data were acquired strictly according to operation manual. MatriXX was calibrated at each QA session to remove drift in linac output. Of course, routine QA procedure was rigorously performed prior to every COMPASS measurement. In addition, stability of linac delivery was verified at each QA session using a standard head and neck (H&N) RapidArc treatment plan.

MLC Error Introduction

MLC errors mainly include individual MLC positional errors, random MLC positional errors and systematic positional errors. Individual MLC positional errors or random MLC positional errors have little dosimetric effect on RapidArc plans (24). As such, systematic MLC positional errors were introduced in this study. An in-house Python program based on Pydicom (version 2.1.2) was applied to introduce systematic MLC errors by manipulation of DICOM RT files. In total, 20 plans were randomly selected from enrolled RapidArc plans. As described in **Figure 2**, the introduced systematic MLC errors included increase and decrease of the distance between leaf pairs in beam field. The magnitude of systematic MLC errors included $\pm 0.2\text{mm}$, $\pm 0.4\text{mm}$, $\pm 0.6\text{mm}$, $\pm 0.8\text{mm}$, $\pm 1\text{mm}$, $\pm 2\text{mm}$, $\pm 3\text{mm}$, $\pm 4\text{mm}$, $\pm 5\text{mm}$, respectively. If any magnitude of MLC errors has led to a negative leaf gap in some leaf pairs, the gap of relevant leaf pair would have been set to the minimal dynamic leaf gap of linac. After systematic MLC errors were introduced into the original plans, the modified RT files were re-imported into TPS for dose calculation. 19 plans were generated per patient (1 baseline plan plus 18 different magnitude of MLC error-introduced plans), so there were 380 plans in total. In the same manner, a total of 180 plans, including 162 error-introduced plans and 18 original plans, were also generated as a validation dataset.

Data Analysis

With the measurement-QA results of COMPASS system, the % DE of DVH information between original plans and modified plans was calculated as follows:

$$\%DE = \frac{D_{mod} - D_{ori}}{D_{ori}} \times 100\% \quad (1)$$

%DE refers to the relative dose percentage change between the original plan dose (D_{ori}) and the measured dose from the modified plans (D_{mod}). The extracted DVH metrics included the dose received by 95% of PTV (PTV_{95}), the dose received by 5% of PTV (PTV_5), the mean dose of PTV (PTV_{mean}), the mean dose of bladder ($Bladder_{mean}$), rectum ($Rectum_{mean}$) and femoral heads (FH_{mean}). These QA results were reviewed for statistical analysis. The Kolmogorov-Smirnov test was performed to determine whether a data set of each DVH metric was well-modeled by the normal distribution ($p > 0.05$).

Once the %DE of DVH metrics was collected from the measurement-QA results of 110 RapidArc plans, process-based DVH action levels were obtained by statistical process control. This method was recommended in AAPM TG-218 report, and the general formula is as follows:

$$\begin{cases} AL = T \pm \beta \sqrt{\sigma^2 + (x - T)^2} \\ TL = x \pm 2.660 \cdot \overline{mR} \end{cases} \quad (2)$$

AL is the action limits. T refers to the process target value; σ^2 and x indicate the process variance and process mean respectively. As a constant, β is set to 6.0 in this equation. TL is the tolerance limits. Calculated by equation $\overline{mR} = \frac{1}{n-1} \sum_{i=2}^n |x_i - x_{i-1}|$, \overline{mR} indicate the moving range. n is the total number of measurements. The details were explicitly explained in AAPM TG-218 report (13).

A linear regression analysis was implemented between the %DE of DVH metrics and introduced MLC errors. By linear regression analysis, the slope, indicating the dose percentage change per mm of MLC error in each DVH metric, was checked. In the study, all presented DVH-based action levels were investigated by linear regression model. The detectability of DVH-based action levels was evaluated through the theoretical detection limit in catching

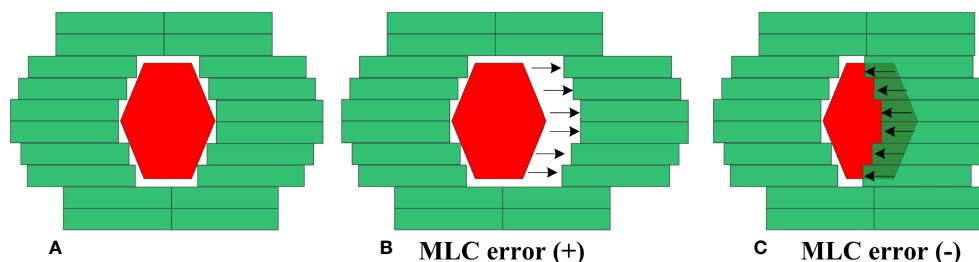


FIGURE 2 | Creation of MLC error-induced RapidArc plan. (A) original plan (B) error-induced plan with positive MLC errors (C) error-induced plan with negative MLC errors.

system MLC errors. The optimal DVH-based action levels in different DVH metrics were also examined by 180 test RapidArc plans. The error detection rate of DVH-based action levels was compared in different DVH metrics. The original plans were considered to be error-free plans. Therefore, the error detection rate of original plans represented the false error detection rate of the evaluated DVH action levels. The error detection rate of error-introduced plans represented the detectability of evaluated DVH-based action levels in identifying systematic MLC errors.

RESULTS

QA Results and Process-Based DVH Action Levels

COMPASS measurement-based QA was implemented for all enrolled RapidArc plans, and QA results were acceptable (see **Table 1**). The average GPRs (3%/2mm, 10% dose threshold) were $99.23\% \pm 0.58\%$, and the mean GI (3%/2mm, 50% dose threshold) were 0.38 ± 0.04 . Meanwhile, the linear accelerator was calibrated in every routine QA procedure to keep its output fluctuation within $\pm 1\%$, and verification of the standard H&N RapidArc plan generated an average GI 0.44 ± 0.06 , indicating a stable RapidArc delivery by linac during data collection.

As described in **Table 2**, the %DE of DVH metrics in enrolled plans, which were selected to calculate process-based action levels, were within 3%. The Kolmogorov-Smirnov (K-S) test indicated that the %DE in all DVH metrics was in non-normal distribution except Bladder_{mean} and FH_{mean}. Based on the QA results of 110 RapidArc plans, the DVH action levels were acquired, including action limits and tolerance limits. A summary of DVH-based action levels (AL and TL) in different DVH metrics was presented in **Table 2** and **Figure 3**.

MLC Error Sensitivity With Different Action Levels

20 RapidArc plans were selected from the rest of plans, and systematic MLC errors of $\pm 0.2\text{mm}$, $\pm 0.4\text{mm}$, $\pm 0.6\text{mm}$, $\pm 0.8\text{mm}$, $\pm 1\text{mm}$, $\pm 2\text{mm}$, $\pm 3\text{mm}$, $\pm 4\text{mm}$, $\pm 5\text{mm}$ were introduced in these plans. The %DE of DVH metrics were extracted from all the modified plans and original plans. Linear regression analysis was performed by having %DE of DVH metrics as function of the MLC error magnitude. The slope, y-intercept, 95% confidence interval and R^2 were presented in **Table 3**. As shown in **Table 3** and **Figure 4**, the %DE of DVH metrics was found to have a linear dependence on the magnitude of systematic MLC errors when R^2 was greater than 0.9. In addition, from the slope, FH_{mean} was found to be the most sensitive one to the MLC errors in all DVH metrics,

and we also observed that PTV₉₅ tended to be more sensitive to systematic MLC errors than other PTV DVH metrics.

All DVH-based action levels were input into linear regression model to calculate the theoretical detection limits. The detection limits represented the detectability of DVH-based action levels in catching system MLC errors. The smaller the detection limit, the better the detectability. As shown in **Figure 5**, the detection limits of tolerance limits were proven to be better than other DVH-based action levels. Meanwhile, PTV₉₅ and PTV_{mean} had a superior detectability compared with other DVH metrics. As described in **Table 2** and **Figure 5**, the TL of PTV₉₅ ($[-1.54\%, 1.51\%]$) could catch the systematic MLC errors between -0.94 mm and 0.88 mm , and the TL of PTV_{mean} ($[-2.06\%, 0.38\%]$) could also catch the systematic MLC errors between -1.00 mm and 0.80 mm . They were able to detect the systematic MLC errors greater than 1mm . It was worth noting that PTV₉₅ action limits also showed a comparable detectability.

Processed-Based DVH Action Levels Validation in Test Plans

Given that DVH-based tolerance limits had superior detection limits in identifying systematic MLC errors by linear regression model, we applied independent test plans to investigate whether the tolerance limits in different DVH metrics would effectively detect abnormal MLC delivery. Systematic MLC errors ranging from 0.2mm to 5mm were introduced in 18 RapidArc plans, and then 180 RapidArc plans (18 original plans plus 162 error-introduced plans) were generated to be evaluated by tolerance limits of different DVH metrics. The results were shown in **Figure 6** and **Table 4**. For original plans (0 mm MLC error), FH_{mean} showed the maximum error detection rate (27.8%), while PTV₉₅ and PTV₅ showed the minimum error detection rate (5.56%). For the error-introduced plans with systematic MLC errors less than 1mm , the tolerance limits of all DVH metrics were not able to identify systematic MLC errors effectively. FH_{mean} showed the maximum error detection rate of 45.83%, closely followed by PTV_{mean} (26.39%). For the error-introduced plans that systematic MLC errors were more than 1mm , PTV_{mean} had the maximum error detection rate of 88.89%. Of course, the error detection rate up to 94% demonstrated that the tolerance limits in all DVH metric were accurate to identify the error-introduced plans with systematic MLC errors more than 2mm .

DISCUSSION

In this study, a comprehensive and systematic evaluation on DVH-based action levels was performed to cervical cancer

TABLE 1 | Summary of the descriptive statistics for QA results.

Group	Item	N	Mean \pm SD
All enrolled plans	GPRs (3%/2mm, 10%)	148	$99.23\% \pm 0.58\%$
	GI (3%/2mm, 50%)		0.38 ± 0.04
Standard H&N plan	GI (3%/2mm, 50%)	29	0.44 ± 0.06

TABLE 2 | A list of DVH action levels obtained by statistical process control for each structure in RapidArc cervical cancer plans.

Structure	Metric	Mean \pm SD (%)	K-S TEST	Action limits (%) [lower, upper]	Tolerance limits (%) [lower, upper]
PTV	D ₅	-1.53 \pm 0.77	$p=0.002$	[-6.56, 3.57]	[-3.65, 0.66]
	D ₉₅	-0.01 \pm 0.67	$p<0.001$	[-2.08, 2.07]	[-1.53, 1.51]
	D _{mean}	-0.84 \pm 0.53	$p=0.006$	[-3.85, 2.16]	[-2.06, 0.38]
Bladder	D _{mean}	-2.04 \pm 0.83	$p=0.011$	[-8.76, 4.63]	[-4.00, -0.14]
Rectum	D _{mean}	1.73 \pm 1.13	$p=0.200$	[-4.44, 7.82]	[-1.59, 4.97]
Femoral Head	D _{mean}	-0.69 \pm 0.99	$p=0.056$	[-4.51, 3.00]	[-3.21, 1.71]

The bold values indicate that the corresponding data are in non-normal distribution ($p < 0.05$).

RapidArc plans. The results demonstrated that process-based DVH action levels, especially tolerance limits, were more powerful than commonly-used DVH action levels in detecting systematic MLC errors in the patient-specific QA. With a long process, patient-specific QA includes random variations and systematic variations. Random variations refer to unavoidable random fluctuations in the process, and they should be monitored to confirm whether the process is under control (25). Arising from unpredictable, non-random events beyond the expected variability, systematic variations can be avoided (26). As an application of statistical techniques, SPC is able to distinguish random variations from systematic variations by specific control limits (also known as tolerance limits). If the data fall outside the control limits, it indicates that systematic errors may have been introduced in the process. That's why tolerance limits have a greater ability to detect systematic MLC errors in patient-specific QA.

Actually, SPC has been proven to be an effective method to monitor and improve QA process using control charts for a long time (20, 21, 23, 26). A control chart typically consists of an upper control limit, a lower control limit and data points. Control limits are key parameters in control charts. As such, setting appropriate

control limits is a critical step. Fuangrod T et al. (19) suggested that control limits should exclude clinically acceptable variability and focus on the detection of gross errors based on "good historical data". Once the initial data contain a great deal of systematic variations, the control limits may be skewed and their ability in detecting systematic errors may be affected. For this reason, rigorous criteria should be set for selecting and collecting data. The QA results of enrolled RapidArc plans (**Table 1**) indicated that the average values of GPRs and GI met the criteria for the standard set. It is an effective way to eliminate systematic variations of initial data as much as possible. The initial data could also be affected by other factors, such as daily machine output fluctuations, setup variations and detector response variations. In order to minimize the variability introduced by these factors, routine QA procedure and a standard H&N RapidArc plan delivery was performed in every QA process. The QA results demonstrated that RapidArc plan's delivery, QA measurements and equipment all remained in a stable state during data collection.

In addition, data characterization is a very important step in SPC charts. As shown in **Table 4**, the error detection rate of FH_{mean} in error-free plans reached as high as 27.78%. One reason was that the rejected plans did contain systematic errors. As shown in

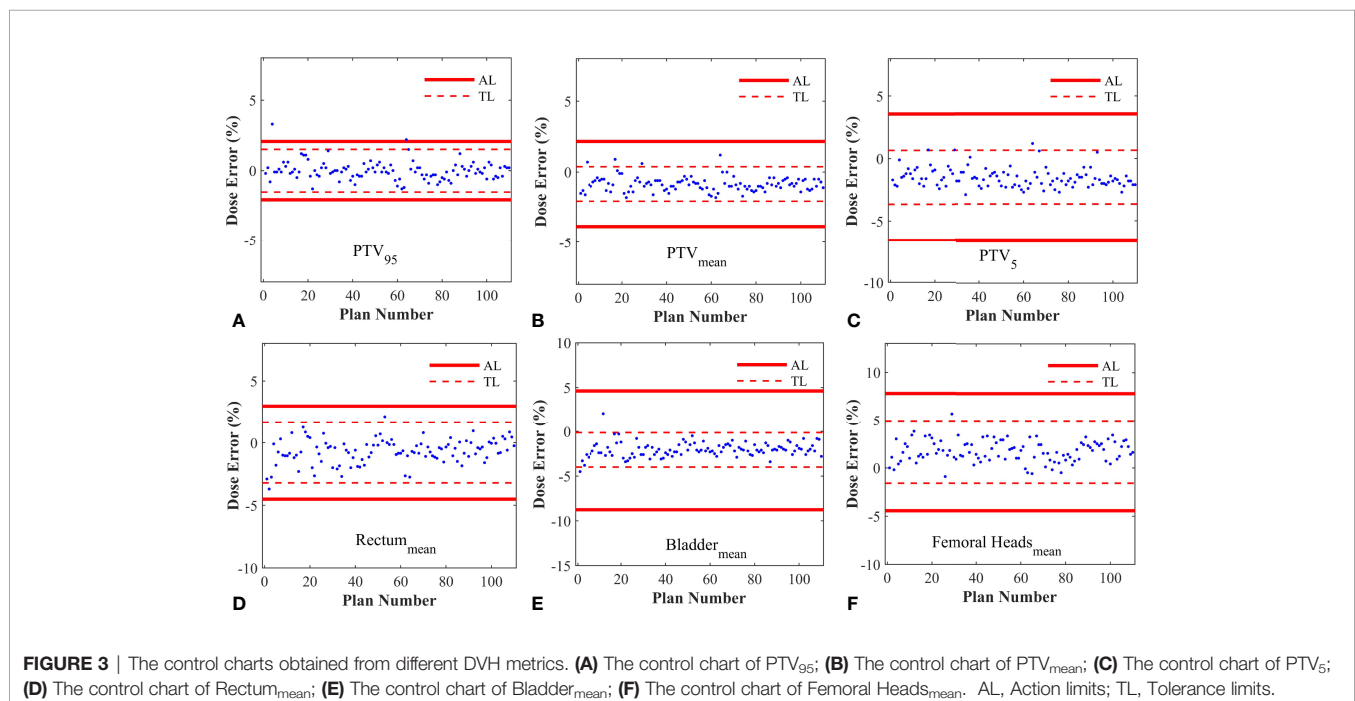


TABLE 3 | A list of the values obtained by linear regression for each structure.

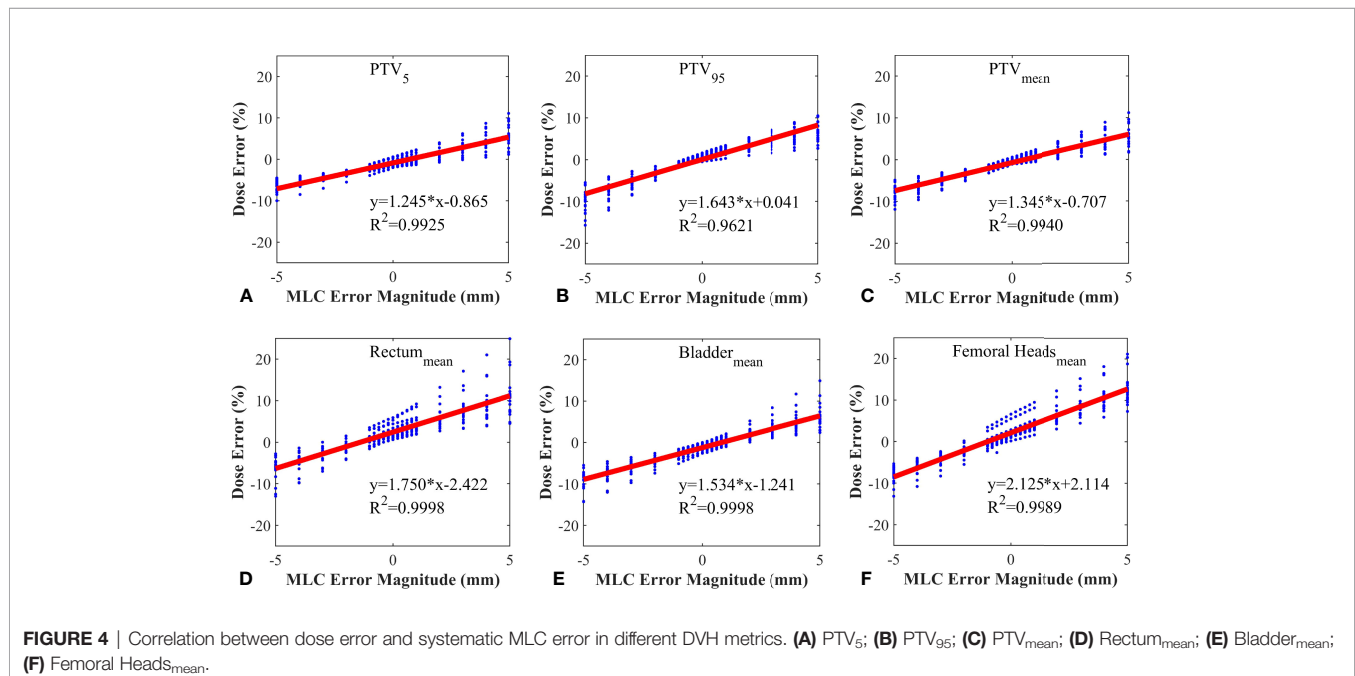
Structure	Metric	Slope(%/mm)	y-intercept(%)	Lower 95% CI(%/mm)	Upper 95% CI(%/mm)	R ²
PTV	D ₅	1.245	-0.865	1.191	1.299	0.9929
	D ₉₅	1.643	0.041	1.476	1.810	0.9621
	D _{mean}	1.345	-0.707	1.291	1.398	0.9940
Bladder	D _{mean}	1.534	-1.241	1.522	1.545	0.9998
Rectum	D _{mean}	1.750	2.422	1.736	1.763	0.9998
Femoral head	D _{mean}	2.125	2.114	2.098	2.161	0.9989

Figure 5, the detection limits in the same DVH metric were extremely asymmetrical in two directions and shifted from zero. M. Steers et al. (17) concluded that the shifts in different directions may indicate different systematic errors caused by algorithm model, machine or other devices. The results of our study were also inevitably influenced by the deviation between AAA and CCC algorithm, the resolution limit of detector array, setup variations and other factors. That was why AAPM TG-218 highlighted that the systematic errors should be eliminated to the degree possible during the QA process. The other reason was that the DVH metrics were not all normally distributed (**Table 2**). Xiao et al. (18) have demonstrated that using normal-based control charts may result in incorrect decisions when the initial data is non-normally distributed. They have concluded that applying a normal-based control chart to a non-normal distribution process could increase the type risk and false alarm rate. The solution was to either transform the non-normal data into normal data, or to use a non-normal-based method to obtain the control chart.

After filtering the data and characterizing the data distribution, the control limits (or tolerance limits) were established by repeating the “Identify-Eliminated-Recalculate” procedure several times (18). The QA results within tolerance limits implied that the process was only affected by random

errors and considered to be under control. As long as the process is under control, action limits are also calculated from QA measurement results over a time period. As shown in **Table 2**, action limits in each DVH metrics were more lenient than tolerance limits. It is because that action limits are set as a minimum level of process performance, and defined as the boundaries outside which a process could cause a negative clinical impact for the patient (13). Therefore, the last step in patient-specific QA process is to compare tolerance limits with action limits, and to ensure that the tolerance limits are within the action limits. Furthermore, it was worth noting that differences in the tolerance limits and action limits were observed between different structures and DVH metrics. It is clear that universally action levels are not adequate for patient-specific QA, especially when individualized evaluation is drawing more and more attention in QA process. Above all, there are variations in QA process due to different factors, but SPC provides a valuable and specific method to identify these variations, ensuring the precision of treatment delivery and safety of patients in radiotherapy.

Gantry, collimator, dose rate and MLC are not only the crucial modulated parameters in RapidArc plans, but also the root cause of variations in QA process. T. Betzel et al. (27) have



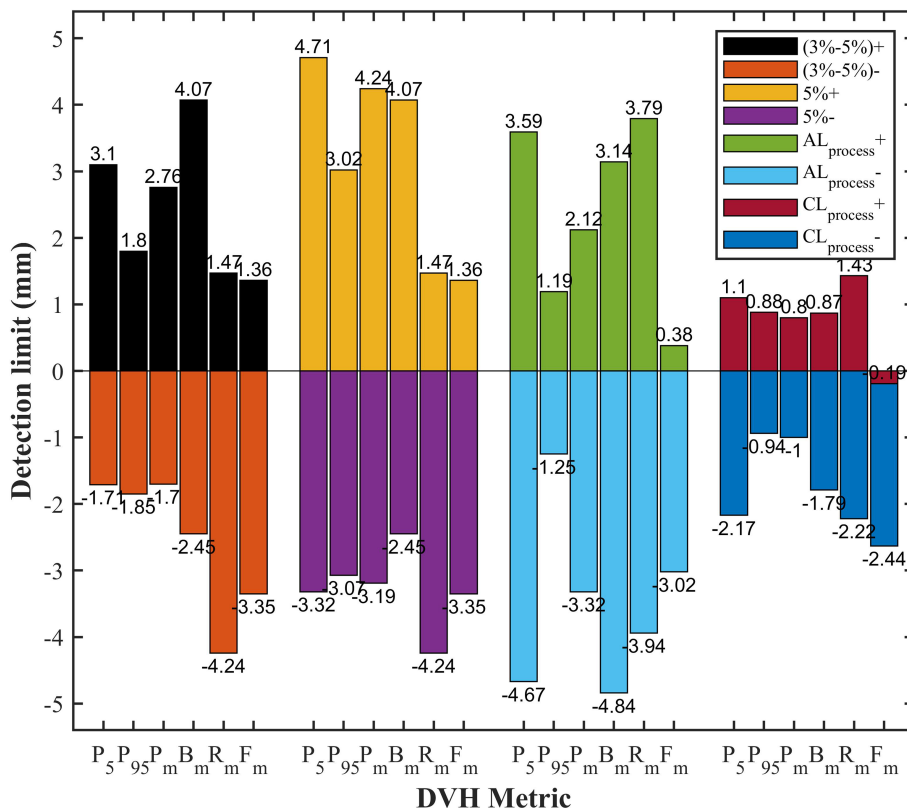


FIGURE 5 | The comparison of detection limits in different DVH-based action levels. P₅, PTV₅; P₉₅, PTV₉₅; P_m, PTV_{mean}; B_m, Bladder_{mean}; R_m, Rectum_{mean}; F_m, Femoral-heads_{mean}; (3%-5%)+, 3% DVH action levels for target volume and 5% DVH action levels for OARs in positive direction; (3%-5)-, 3% DVH action levels for target volume and 5% DVH action levels for OARs in negative direction; 5%+, 5% DVH action levels for target volume in positive direction; 5%-, 5% DVH action levels for target volume in negative direction; AL_{process}+, process-based action limits in positive direction; AL_{process}-, process-based action limits in negative direction; CL_{process}+, process-based tolerance limits in positive direction; CL_{process}-, process-based tolerance limits in negative direction.

demonstrated that RapidArc plans are more susceptible to MLC variations rather than incorrect gantry or collimator angle, and dose rate variations. Moreover, MLC plays a decisive role in the beam modulation, therefore, high-quality radiation therapy requires optimal MLC (28, 29). Small errors in MLC positioning may lead to an adverse consequence to dose distribution, so the accuracy of MLC leaf positions should be monitored carefully. However, a considerable number of studies have reported that the dosimetric impact caused by individual MLC positioning offsets or random MLC errors is relatively insignificant. Wang et al. (24) have also reported that MLC positional errors have a distinguishable dosimetric effect relative to other linac errors. It was not hard to understand why our study focused on the effect of the systematic MLC positional errors in DVH metrics. As shown in **Table 3**, the dose sensitivity of systematic MLC errors was 1.643%/mm, 1.345%/mm for PTV₉₅ and PTV_{mean}, respectively. These results can be compared with that in previous studies (29, 30). The FH_{mean} was found to have the highest dose sensitivity of 2.15%/mm. It may be attributed to the fact that femoral heads were located in low-dose regions, thus the relative difference was magnified. More importantly, we observed a strong positive linear

correlation between DVH metrics and systematic MLC errors when R² was greater than 0.9. It was consistent with values previously reported in other studies (30–32). The strong relationship between DVH metrics and systematic MLC errors implied that reasonable DVH-based action levels could identify and eliminate the systematic MLC errors to the degree possible. However, the comparison of detection limits indicated that commonly used DVH-based action levels were insensitivity enough to detect systematic MLC errors with clinically significant DVH differences. A. Sdrolia et al. (4) have reported that action levels should not be blindly copied, and they must be locally determined based on tumor response and normal tissue complication. So, how to select reasonable DVH-based action levels was one of the issues that should be solved in our study.

It has been suggested in a number of literatures (10, 32, 33) that the deviations in DVH metrics for target volume should be kept within ±2%. The tolerance limits of PTV₉₅ and PTV_{mean} in our study were consistent with this requirement. As described in **Figure 6** and **Table 4**, the comparison of detection limits and error detection rate demonstrated that systematic MLC errors up to 1mm could be effectively caught by tolerance limits. In numerous publications (6, 13, 29, 34), it has been found that

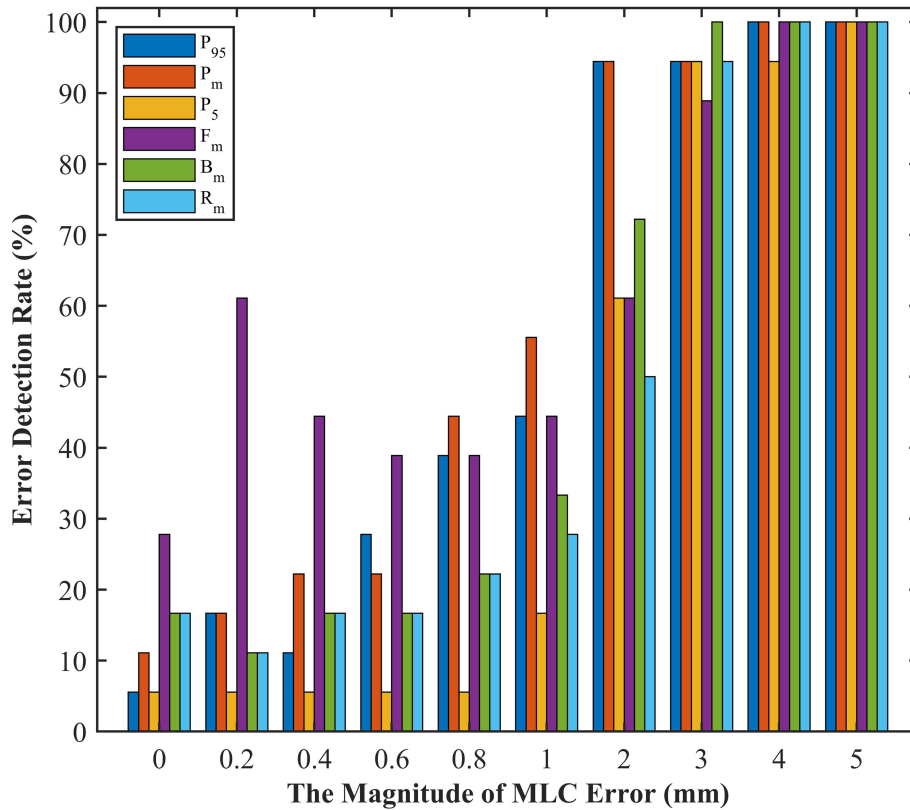


FIGURE 6 | The error detection rate of test plans in different magnitude of systematic MLC errors. P₅, PTV₅; P₉₅, PTV₉₅; P_m, PTV_{mean}; B_m, Bladder_{mean}; R_m, Rectum_{mean}; F_m, Femoral-heads_{mean}.

the systematic MLC errors up to 1mm can produce a clinically relevant influence on the dose distribution, but leaf position errors less than 2mm is unable to be detected by traditional gamma analysis with commonly used criterion of 3%/3mm. It has also been recommended in AAPM TG-142 report that the leaf position accuracy should be within ±1mm (35). Therefore, process-based DVH action levels, especially tolerance limits, can not only detect systematic MLC errors with clinical significance, but also have a superior detectability than commonly used action levels. Notably, the highest error detection rate in detecting plans with systematic MLC errors less than 1mm was only 45.83% (F_m). However, Rangel et al. (36) have suggested that systematic MLC errors need to be limited to 0.3mm. Oliver

et al. (32) have suggested that the systematic MLC errors should be within 0.63mm to keep the PTV95 within 2%. As for the low efficiency in catching minor systematic MLC errors, one factor was that the detectability was limited to the low-spatial resolution array detectors. Woon et al. (37) have also found that a 0.75mm systematic MLC error is undetected due to the poor resolution of array detectors. Another factor was that the sensitivity to MLC errors could vary according the treatment sites, delivery techniques or QA measurement instruments. In addition, the introduced MLC errors in our study were systematic MLC shift errors and tended to have a smaller effect on the DVH metrics than systematic open/closed MLC errors (38). Given the above, the insensitivity of tolerance limits to

TABLE 4 | Error detection rate based on tolerance limits of different DVH metrics in different magnitude of systematic MLC errors.

Error magnitudes (mm)	Error detection rate (%)					
	P ₉₅	P _m	P ₅	B _m	R _m	F _m
0	5.56%	11.11%	5.56%	16.67%	16.67%	27.78%
0.2~0.8	23.61%	26.39%	5.56%	16.67%	16.67%	45.83%
1~5	86.67%	88.89%	73.33%	81.11%	74.44%	78.89%

P₅, PTV₅; P₉₅, PTV₉₅; P_m, PTV_{mean}; B_m, Bladder_{mean}; R_m, Rectum_{mean}; F_m, Femoral-heads_{mean}.
The bold values indicate the best error detection rate in different magnitude of systematic MLC errors.

detect systematic MLC errors below 1mm was reasonable, and process-based DVH action levels may offer reasonable action levels to identify and catch systematic MLC errors with clinical significance.

Once the method for obtaining DVH-based action levels was determined, the next challenge was to select appropriate DVH metrics from numerous evaluation structures and make a final QA determination. It is widely known that MLC positional errors can lead to clinically significant dose deviation in the PTV and adjacent OARs. Many studies (1, 4, 32, 39) have demonstrated that DVH metrics of PTV are the primary concern with the quality and deliverability of VMAT plans. The dose distribution in the PTV is more sensitive to MLC positional errors because that PTV is mostly deep-seated and close to isocenter (40), and MLC apertures varies dynamically along with the shape of PTV. As described in **Table 4**, PTV₉₅ showed a higher error detection rate in plans with systematic MLC errors greater than 1mm, but a lower error detection rate in original (error-free) plans. This implied that PTV₉₅ may be a representative DVH metric in detecting systematic MLC errors in this study. The change of DVH metrics in OARs depended on the location of OARs relative to the target volume and the size of OARs. The dose sensitivity of OARs in this study was in agreement with the study of Nithiyantham et al. (29) that the OARs partially located within target volume or having a smaller volume showed an obvious dose difference in DVH metrics with MLC positional errors. However, our previous study has found that the OARs with smaller volume or far away from the isocenter could generate false positive results. Therefore, after completing the evaluation of target volume, it is necessary to carry out an objective analysis to the dose difference of OARs in DVH metrics.

Of course, there are some limitations in this study. First of all, the MLC positional errors included in this study were oversimplified. MLC positional errors were correlated with the MLC speed, gantry sag and other factors. In the future, more machine errors need to be simulated and introduced to investigate the dose sensitivity to these errors. Secondly, the data of DVH metrics were not all normally distributed. The non-normal distribution may result in incorrect decisions for DVH-based action levels. Further research is needed to address issue of non-normal distribution so as to obtain more reliable action levels by SPC. Finally, our findings may be confined to our devices, treatment sites and sample size. However, process-based DVH

action levels could be available to other delivery techniques, treatment sites or QA tools. A large sample size is required and further investigation need to be implemented for other dosimetric system or treatment sites.

CONCLUSIONS

In conclusion, for the cervical cancer RapidArc plans, through tolerance limits based on statistical process control, reasonable DVH-based action levels can be acquired to identify and catch systematic MLC errors more than 1mm in 3D dose verification. During the evaluation of DVH metrics, a comprehensive analysis focusing on target volume should be implemented on structure by structure in order to ensure the quality and deliverability of radiotherapy plans. Although the OARs in low-dose regions showed a relatively stronger dose sensitivity to systematic MLC errors, their process-based DVH action levels may be accompanied with higher false error detection rate.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

HZ and WL have contributed equally to this work. XY, HZ and WL: drafting of work, analysis and interpretation of trials and literature, drafting of manuscript, and manuscript review. HZ, XY collected the data, reviewed the literature, and wrote the paper. HC and YL prepared the figure and contributed in the revision of the literature. All authors contributed to the article and approved the submitted version.

FUNDING

This work was sponsored by Natural Science Foundation of Chongqing, China (cstc2021jcyj-msxmX0138).

REFERENCES

- Shen L, Chen S, Zhu X, Han C, Zheng X, Deng Z, et al. Multidimensional Correlation Among Plan Complexity, Quality and Deliverability Parameters for Volumetric-Modulated Arc Therapy Using Canonical Correlation Analysis. *J Radiat Res* (2018) 59(2):207–15. doi: 10.1093/jrr/rrx100
- Li J, Wang L, Zhang X, Liu L, Li J, Chan MF, et al. Machine Learning for Patient-Specific Quality Assurance of VMAT: Prediction and Classification Accuracy. *Int J Radiat Oncol Biol Phys* (2019) 105(4):893–902. doi: 10.1016/j.ijrobp.2019.07.049
- Otto K. Volumetric Modulated Arc Therapy: IMRT in a Single Gantry Arc. *Med Phys* (2008) 35(1):310–7. doi: 10.1118/1.2818738
- Sdrolia A, Brownsword KM, Marsden JE, Alty KT, Moore CS, Beavis AW. Retrospective Review of Locally Set Tolerances for VMAT Prostate Patient Specific QA Using the COMPASS(R) System. *Phys Med* (2015) 31(7):792–7. doi: 10.1016/j.ejmp.2015.03.017
- Mancuso GM, Fontenot JD, Gibbons JP, Parker BC. Comparison of Action Levels for Patient-Specific Quality Assurance of Intensity Modulated Radiation Therapy and Volumetric Modulated Arc Therapy Treatments. *Med Phys* (2012) 39(7):4378–85. doi: 10.1118/1.4729738
- Wootton LS, Nyflot MJ, Chaovalitwongse WA, Ford E. Error Detection in Intensity-Modulated Radiation Therapy Quality Assurance Using Radiomic Analysis of Gamma Distributions. *Int J Radiat Oncol Biol Phys* (2018) 102(1):219–28. doi: 10.1016/j.ijrobp.2018.05.033

7. Nelms BE, Zhen H, Tome WA. Per-Beam, Planar IMRT QA Passing Rates do Not Predict Clinically Relevant Patient Dose Errors. *Med Phys* (2011) 38 (2):1037–44. doi: 10.1118/1.3544657
8. Zhen H, Nelms BE, Tome WA. Moving From Gamma Passing Rates to Patient DVH-Based QA Metrics in Pretreatment Dose Qa. *Med Phys* (2011) 38(10):5477–89. doi: 10.1118/1.3633904
9. Coleman L, Skourou C. Sensitivity of Volumetric Modulated Arc Therapy Patient Specific QA Results to Multileaf Collimator Errors and Correlation to Dose Volume Histogram Based Metrics. *Med Phys* (2013) 40(11):111715. doi: 10.1118/1.4824433
10. Alharthi T, Pogson EM, Arumugam S, Holloway L, Thwaites D. Pre-Treatment Verification of Lung SBRT VMAT Plans With Delivery Errors: Toward a Better Understanding of the Gamma Index Analysis. *Phys Med* (2018) 49:119–28. doi: 10.1016/j.ejmp.2018.04.005
11. Gay SS, Netherton TJ, Cardenas CE, Ger RB, Balter PA, Dong L, et al. Dosimetric Impact and Detectability of Multi-Leaf Collimator Positioning Errors on Varian Halcyon. *J Appl Clin Med Phys* (2019) 20(8):47–55. doi: 10.1002/acm2.12677
12. Ezzell GA, Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT Commissioning: Multiple Institution Planning and Dosimetry Comparisons, a Report From AAPM Task Group 119. *Med Phys* (2009) 36 (11):5359–73. doi: 10.1118/1.3238104
13. Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance Limits and Methodologies for IMRT Measurement-Based Verification QA: Recommendations of AAPM Task Group No. 218. *Med Phys* (2018) 45(4):e53–83. doi: 10.1002/mp.12810
14. Mu G, Ludlum E, Xia P. Impact of MLC Leaf Position Errors on Simple and Complex IMRT Plans for Head and Neck Cancer. *Phys Med Biol* (2008) 53 (1):77–88. doi: 10.1088/0031-9155/53/1/005
15. Visser R, Wauben DJ, de Groot M, Godart J, Langendijk JA, van't Veld AA, et al. Evaluation of DVH-Based Treatment Plan Verification in Addition to Gamma Passing Rates for Head and Neck IMRT. *Radiother Oncol* (2014) 112 (3):389–95. doi: 10.1016/j.radonc.2014.08.002
16. Yi X, Lu WL, Dang J, Huang W, Cui HX, Wu WC, et al. A Comprehensive and Clinical-Oriented Evaluation Criteria Based on DVH Information and Gamma Passing Rates Analysis for IMRT Plan 3d Verification. *J Appl Clin Med Phys* (2020) 21(8):47–55. doi: 10.1002/acm2.12910
17. Steers JM, Fraass BA. IMRT QA: Selecting Gamma Criteria Based on Error Detection Sensitivity. *Med Phys* (2016) 43(4):1982. doi: 10.1118/1.4943953
18. Xiao Q, Bai S, Li G, Yang K, Bai L, Li Z, et al. Statistical Process Control and Process Capability Analysis for Non-Normal Volumetric Modulated Arc Therapy Patient-Specific Quality Assurance Processes. *Med Phys* (2020) 47(10):4694–702. doi: 10.1002/mp.14399
19. Fuangrod T, Greer PB, Simpson J, Zwan BJ, Middleton RH. A Method for Evaluating Treatment Quality Using *In Vivo* EPID Dosimetry and Statistical Process Control in Radiation Therapy. *Int J Health Care Qual Assur* (2017) 30 (2):90–102. doi: 10.1108/IJHCQA-03-2016-0028
20. Pawlicki T, Whitaker M, Boyer AL. Statistical Process Control for Radiotherapy Quality Assurance. *Med Phys* (2005) 32(9):2777–86. doi: 10.1118/1.2001209
21. Pawlicki T, Yoo S, Court LE, McMillan SK, Rice RK, Russell JD, et al. Process Control Analysis of IMRT QA: Implications for Clinical Trials. *Phys Med Biol* (2008) 53(18):5193–205. doi: 10.1088/0031-9155/53/18/023
22. Bellec J, Delaby N, Jouyaux F, Perdrieux M, Bouvier J, Sorel S, et al. Plan Delivery Quality Assurance for CyberKnife: Statistical Process Control Analysis of 350 Film-Based Patient-Specific QAs. *Phys Med* (2017) 39:50–8. doi: 10.1016/j.ejmp.2017.06.016
23. Wang H, Xue J, Chen T, Qu T, Barbee D, Tam M, et al. Adaptive Radiotherapy Based on Statistical Process Control for Oropharyngeal Cancer. *J Appl Clin Med Phys* (2020) 21(9):171–7. doi: 10.1002/acm2.12993
24. Wang Y, Pang X, Feng L, Wang H, Bai Y. Correlation Between Gamma Passing Rate and Complexity of IMRT Plan Due to MLC Position Errors. *Phys Med* (2018) 47:112–20. doi: 10.1016/j.ejmp.2018.03.003
25. Palaniswaamy G, Scott Brame R, Yaddanapudi S, Rangaraj D, Mutic S. A Statistical Approach to IMRT Patient-Specific Qa. *Med Phys* (2012) 39 (12):7560–70. doi: 10.1118/1.4768161
26. Nordström F, af Wetterstedt S, Johnsson S, Ceberg C, Bäck SJ. Control Chart Analysis of Data From a Multicenter Monitor Unit Verification Study. *Radiother Oncol* (2012) 102(3):364–70. doi: 10.1016/j.radonc.2011.11.016
27. Betzel GT, Yi BY, Niu Y, Yu CX. Is RapidArc More Susceptible to Delivery Uncertainties Than Dynamic IMRT? *Med Phys* (2012) 39(10):5882–90. doi: 10.1118/1.4749965
28. Letourneau D, Wang A, Amin MN, Pearce J, McNiven A, Keller H, et al. Multileaf Collimator Performance Monitoring and Improvement Using Semiautomated Quality Control Testing and Statistical Process Control. *Med Phys* (2014) 41(12):121713. doi: 10.1118/1.4901520
29. Nithiyantham K, Mani GK, Subramani V, Mueller L, Palaniappan KK, Kataria T. Analysis of Direct Clinical Consequences of MLC Positional Errors in Volumetric-Modulated Arc Therapy Using 3D Dosimetry System. *J Appl Clin Med Phys* (2015) 16(5):296–305. doi: 10.1120/jacmp.v16i5.5515
30. Kadoya N, Saito M, Ogasawara M, Fujita Y, Ito K, Sato K, et al. Evaluation of Patient DVH-Based QA Metrics for Prostate VMAT: Correlation Between Accuracy of Estimated 3D Patient Dose and Magnitude of MLC Misalignment. *J Appl Clin Med Phys* (2015) 16(3):5251. doi: 10.1120/jacmp.v16i3.5251
31. Oliver M, Bush K, Zavgorodni S, Ansbacher W, Beckham WA. Understanding the Impact of RapidArc Therapy Delivery Errors for Prostate Cancer. *J Appl Clin Med Phys* (2011) 12(3):3409. doi: 10.1120/jacmp.v12i3.3409
32. 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):554–60. doi: 10.1016/j.radonc.2010.06.013
33. Heilemann G, Poppe B, Laub W. On the Sensitivity of Common Gamma-Index Evaluation Methods to MLC Misalignments in Rapidarc Quality Assurance. *Med Phys* (2013) 40(3):031702. doi: 10.1118/1.4789580
34. Stambaugh C, Ezzell G. A Clinically Relevant IMRT QA Workflow: Design and Validation. *Med Phys* (2018) 45(4):1391–9. doi: 10.1002/mp.12838
35. Klein EE, Hanley J, Bayouth J, Yin FF, Simon W, Dresser S, et al. Task Group 142 Report: Quality Assurance of Medical Accelerators. *Med Phys* (2009) 36 (9):4197–212. doi: 10.1118/1.3190392
36. Rangel A, Palte G, Dunscombe P. The Sensitivity of Patient Specific IMRT QC to Systematic MLC Leaf Bank Offset Errors. *Med Phys* (2010) 37(7):3862–7. doi: 10.1118/1.3453576
37. Woon W, Ravindran PB, Ekayanake P, S V, Lim YY, Khalid J. A Study on the Effect of Detector Resolution on Gamma Index Passing Rate for VMAT and IMRT QA. *J Appl Clin Med Phys* (2018) 19(2):230–48. doi: 10.1002/acm2.12285
38. Saito M, Sano N, Shibata Y, Kuriyama K, Komiyama T, Marino K, et al. Comparison of MLC Error Sensitivity of Various Commercial Devices for VMAT Pre-Treatment Quality Assurance. *J Appl Clin Med Phys* (2018) 19 (3):87–93. doi: 10.1002/acm2.12288
39. Bojcheko C, Ford EC. Quantifying the Performance of *In Vivo* Portal Dosimetry in Detecting Four Types of Treatment Parameter Variations. *Med Phys* (2015) 42(12):6912–8. doi: 10.1118/1.4935093
40. Liang B, Liu B, Zhou F, Yin FF, Wu Q. Comparisons of Volumetric Modulated Arc Therapy (VMAT) Quality Assurance (QA) Systems: Sensitivity Analysis to Machine Errors. *Radiat Oncol* (2016) 11(1):146. doi: 10.1186/s13014-016-0725-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Lu, Cui, Li and Yi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.