Journal of Radiotherapy in Practice

cambridge.org/jrp

Wobbling nature of gamma passing rate as a function of calibration field sizes in patient-specific quality assurance

Sathiya Raj[1], Venugopal Sundaram[2], Henry Finlay Godson[3] and Retna John[3]

[1]Department of Radiation Physics, Kidwai Memorial Institute of Oncology, Dr. M.H. Marigowda Road, Bengaluru, India; [2]Meherbai Tata Memorial Hospital, R53J+3HJ, Stocking Road, Northern Town, Bistupur, Jamshedpur, Jharkhand, India and [3]Department of Radiation Oncology, Christian Medical College Vellore, Ranipet Campus, Kilminnal Village, Ranipet District, TN, India

Original Article

Cite this article: Raj S, Sundaram V, Godson HF, and John R. (2024) Wobbling nature of gamma passing rate as a function of calibration field sizes in patient-specific quality assurance. *Journal of Radiotherapy in Practice*. **23**(e2), 1–6. doi: 10.1017/S1460396923000444

Received: 18 March 2023
Revised: 7 December 2023
Accepted: 13 December 2023

Keywords:
Gamma analysis; patient-specific QA; calibration; IMRT

Corresponding author:
Sathiya Raj;
Email: sathiarajmedphy@gmail.com

Abstract

Purpose: This study aimed to investigate the influence of calibration field size on the gamma passing rate (GPR) in patient-specific quality assurance (PSQA).
Methods: Two independent detectors, PTW OCTAVIUS 4D (4DOCT) and Arc Check, were utilised in volumetric modulated arc therapy plans for 26 patients (14 with Arc Check and 12 with 4DOCT). Plans were administered using Varian Unique machine (with 4DOCT) and Varian TrueBeam (with Arc Check), each employing different calibration factors (CFs): 4×4, 6×6, 8×8, 10×10, 12×12 and 15×15 cm^2 field sizes. Gamma analysis was conducted with 2%2mm, 2%3mm and 3%3mm gamma criteria.
Results: GPR exhibited variations across different CFs. GPR demonstrated an increasing trend below 10×10 cm^2 CFs, while it displayed a decreasing trend above 10×10 cm^2. Both detectors exhibited similar GPR patterns. The correlation between 4DOCT and Arc Check was strong in tighter criteria (2%2mm) with an R^2 value of 0·9957, moderate criteria (2%3mm) with an R^2 value of 0·9868, but reduced in liberal criteria (3%3mm) with an R^2 value of 0·4226.
Conclusion: This study demonstrates that calibration field sizes significantly influence GPR in PSQA. This study recommends the plan specific calibration field must obtain to calibrate the QA devices for modulated plans.

Introduction

Patient-apecific quality assurance (PSQA) is an indispensable component in the radiotherapy treatment chain. PSQA can be defined as a pre-treatment process involving measurements to confirm the deliverability and quality of the patient's treatment plan.[1] This ensures that the planned radiation dose aligns with the intended treatment and meets the required quality standards before the actual patient treatment begins. Various factors, such as the wrong position of the multi-leaf collimator (MLC), MLC speed and fluctuations in the linear accelerator (LINAC) output, beam stability, and planning grid size, may contribute to errors in treatment delivery.[2] These errors can lead to deviations from the planned dose, necessitating identification before treating patients. To identify errors in treatment delivery, PSQA must be performed. While simple ion chamber measurements can provide one-dimensional (1D) information, it may not be sufficient for identifying errors in complex volumetric-modulated arc therapy (VMAT) and intensity-modulated radiation therapy (IMRT) plans. To obtain two-dimensional (2D) or three-dimensional (3D) dose information, many commercial array-based detectors are available,[3–7] equipped with built-in analyzing software. The planned dose and measured dose can be compared using the gamma index, quantifying the gamma passing rate (GPR) in percentage. The gamma analysis method described by Low et al. (1998)[8] is commonly used, and most commercial array-based detectors have the capability to calculate the GPR. Before commencing PSQA, the 2D array must be calibrated with a conventional field size,[9] for example, 10×10 cm^2 or 5×5 cm^2. Calibration is performed by delivering a known dose and measuring the actual dose, resulting in a calibration factor (CF) that is then applied to all detectors embedded in the 2D array. VMAT/IMRT plans are constructed using multiple segments, and the dimensions of these segments may not be equal to the calibration field size. This discrepancy could introduce uncertainty in the GPR of the given plan. This forms the hypothesis of our study, and its graphical representation is clearly shown in Figure 1. Wei Luo et al. (2018) reported that conventional field size-based calibration may not yield accurate results.[10] They suggested that plan-specific calibration would provide more accurate results, especially for complex plans.[10] However, the reasons behind the changes in the GPR for different calibration field sizes were not addressed in their study, and the results were not compared with another

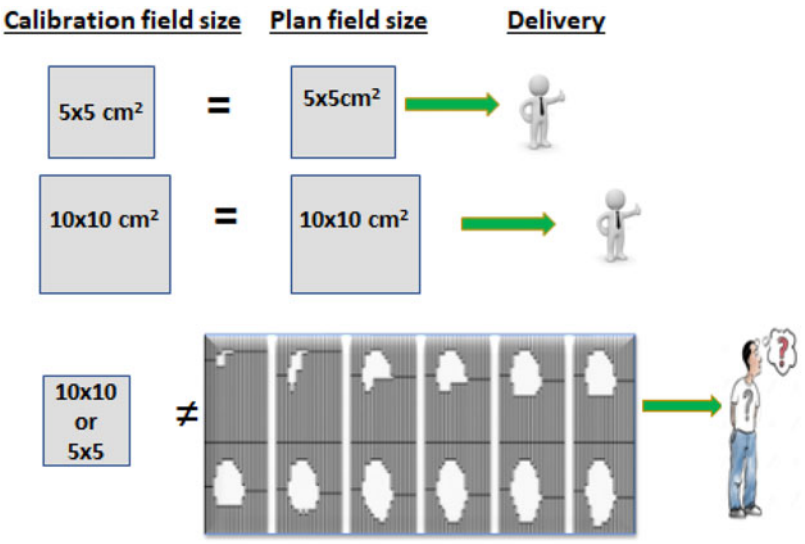

Figure 1. Simple calibration square field not true representation of the true complex segments of the VMT/IMRT plans

dosimeter. In this study, our intention is to identify the reasons for changes in GPR as a function of calibration field size and to investigate the results using two different dosimeters.

Materials and Methods

The study was conducted at two distinct radiotherapy centers. In the first center, PSQA was carried out using the PTW Octavius 4D detector (4DOCT), and the treatment plans were delivered by the Varian Unique linear accelerator. At the second centre, PSQA procedures were performed utilising the Arc Check detector, and the treatment plans were delivered using the Varian True Beam linear accelerator. Different patient samples were used for each centre, and a total of 26 plans were delivered. It is important to note that the same patient plans were not employed for both detectors. The primary focus of this study is not to compare the performance of the 4DOCT and Arc Check detectors.

4DOCT

A total of 12 patient plans were retrospectively selected for this study. VMAT plans were generated for all patients using the Eclipse Treatment Planning System (TPS). Measurements were conducted on a Varian Unique-Performance linear accelerator, utilising 6MV X-ray photons. For dosimetric analysis, 4DOCT phantom, in conjunction with a vented ion chamber array (PTW 729 array), was employed. The 2d array is integrated with 729 ion chambers distributed over a 27×27 cm^2 area. Each chamber has a volume of 0·125 cc, with a center-to-center distance of 1 cm.[9] The array can be inserted into the 4D phantom, which moves synchronously with the gantry. An inclinometer, acting as a gantry angle sensor, is positioned in the stem of the gantry to facilitate this movement. Since the incident photon beam is consistently perpendicular to the detector array, there is no need for an angular dependency correction factor. Data acquired by the 4DOCT were analysed using the software provided by PTW, known as Verisoft –v7·2.

Arc check

A total of 14 patient plans were retrospectively selected for this study. VMAT plans were generated for all patients using the Eclipse TPS. The plans were subsequently delivered using the Varian True beam SVC V2·7 linear accelerator. For the verification of these plans, the Arc Check detector was employed. The Arc Check detector is a cylindrical acrylic phantom equipped with a three-dimensional array consisting of 1386 diode detectors, spaced at 1 cm intervals.[11] Verification plans were created for all patient plans, with the density of polymethyl methacrylate (PMMA) overridden. The phantom was positioned with a couch rotation of zero degrees, and measurements were conducted in the actual plan geometry.

Choice of plan delivery on detectors

AAPM TG 218 outlines various methods for PSQA,[2] including: (a) true composite (TC) delivery, (b) perpendicular field-by-field (PFF) delivery, (c) perpendicular composite (PC) delivery and d) PFF or PC delivery on a 2D array device mounted on the treatment head. In the current study, TC delivery was adopted for Arc Check, as the Arc Check remained stationary, similar to a patient on the treatment couch. During TC delivery, the gantry rotates around the stationary Arc Check. For the 4DOCT measurements, the detector was consistently perpendicular to the incident radiation, making it akin to PC delivery. In both measurement scenarios, the plan was delivered using actual treatment parameters. This encompassed monitor units, gantry angles, collimator settings, couch angles, jaws, and MLC leaf positions.

Gamma analysis

Each plan was calculated and delivered using six CFs (4×4, 6×6, 8×8, 10×10, 12×12, and 15×15 cm^2). For each calibration field size the QA devices were calibrated independently. Consequently, the overall number of delivered plans amounted to 156 (26 plans $\times$ 6 CF = 156). Gamma analysis was conducted for both detectors using 2%2mm, 2%3mm, and 3%3mm gamma criteria, with a 10% low dose threshold in global normalisation mode.

Uncertainty analysis

Type A error in measurements was evaluated by conducting repeated measurements. It's important to note that the data collection spanned multiple sessions, not confined to a single day. Throughout each data collection session, intentional variations in the phantom setup were introduced, and the associated uncertainty

https://doi.org/10.1017/S1460396923000444 Published online by Cambridge University Press

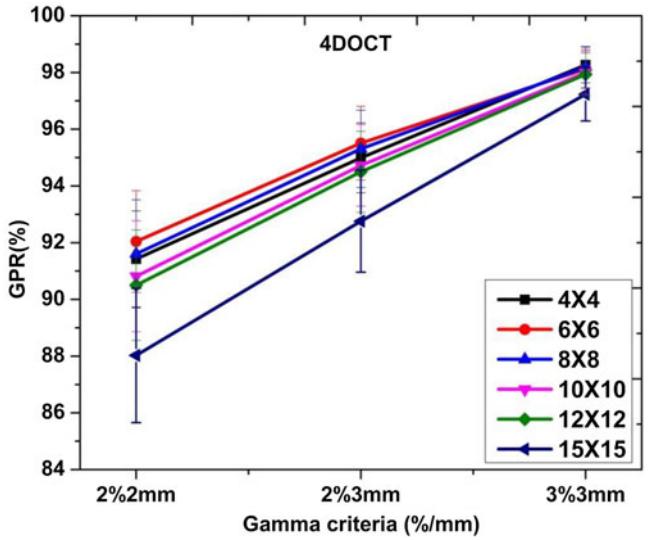

Figure 2. GPR for different calibration field size in 4DOCT

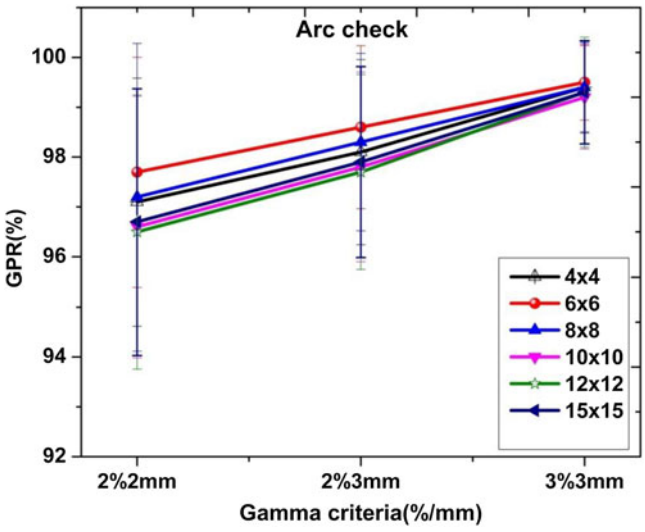

Figure 3. GPR for different calibration field size (Arc Check)

was systematically considered in the study. To quantify this uncertainty, the phantom setup was intentionally disturbed and corrected before plan delivery, a process repeated ten times to calculate the relative uncertainty. Similarly, to address uncertainties linked with plan delivery, the same plan was executed multiple times, and the relative uncertainty was derived from these measurements. Considering that machine output can introduce variations in measurements, this aspect was also incorporated into the overall uncertainty budget. The total uncertainty was calculated using the summation in quadrature method, expressed by the following formula:

$$u_{(c)}(y) = \sqrt{\sum_{i=0}^{n} [ciu(xi)]^2}$$

where, $u_{(c)}(y)$ = combined uncertainty, c_i - sensitivity coefficient, $u(x_i)$ - standard uncertainty

Results

4DOCT

Figure 2 illustrates variations in GPR for different calibration field sizes. For the 2%2mm gamma criterion, the minimum GPR was 88·035% (15 × 15), and the maximum GPR was 92·04% (6 × 6). Under the 2%3mm criterion, the minimum GPR was 92·75% (15 × 15), and the maximum GPR reached 95·5% (6 × 6). For the 3%3mm criterion, the minimum GPR was 97·23% (15 × 15), and the maximum GPR was 98·3% (4 × 4). Conducting a Student-T test revealed statistical significance ($p < 0·0001$) in GPR between 10 × 10 calibrations and 15 × 15 for all gamma criteria. In the 3% 3mm gamma criterion, the GPR for the 8 × 8 calibration field demonstrated statistical significance ($p < 0·05$) compared to the 10 × 10 calibration. Similarly, in the 2%3mm gamma criterion, the GPR significantly differed ($p < 0·05$) from the GPR based on the 10 × 10 calibration.

Arc Check

Variations in the GPR for different calibration field sizes are presented in Figure 3. For the 2%2mm gamma criterion, the GPR ranged from a minimum of 96·6% (10 × 10) to a maximum of 97·7% (6 × 6). Under the 2%3mm criterion, the GPR varied from a minimum of 97·7% (12 × 12) to a maximum of 98·6% (6 × 6). For the 3%3mm criterion, the GPR ranged from a minimum of 99·2% (10 × 10) to a maximum of 99·5% (6 × 6).Conducting a Student-t test revealed statistical significance ($p < 0·0001$) in GPR between 10 × 10 and 4 × 4, 6 × 6 and 8 × 8 for all gamma criteria. Interestingly, for the 15 × 15 calibration field-based GPR, no statistical significance was observed compared to the 10 × 10, irrespective of the gamma criteria. However, for the 12 × 12 calibration field, only the 2%3mm gamma criterion showed statistical significance ($p < 0·05$).

Correlation of 4DOCT and arc check

To assess the response of both dosimeters in terms of the GPR across different CFs, a correlation curve was established. The GPR values for both 4DOCT and Arc Check were plotted, and a second-order polynomial curve was fitted as the best representation of the curve (refer to Figure 4(a), (b), and (c)). In the more stringent criteria of 2%2mm, the R^2 value was 0·9957, indicating a highly robust fit. In the moderate criteria (2%3mm), the R^2 value remained high at 0·9868, demonstrating a strong correlation. However, in the more lenient criteria (3%3mm), the R^2 value decreased to 0·4226 shows that less correlation.

Discussion

Table 1 shows the expected uncertainty of the present study. The overall uncertainty was less than 0·53% in 4DOCT and less than 1·5% in Arc check.

IMRT plans often consist of multiple irregular static or dynamic segments as shown in Figure 1, each with varying dimensions. These segments may not align with the conventional calibration field size of 10 × 10 cm², potentially impacting the beam characteristics. Therefore, it is crucial to analyse the calibration field size dependency of the GPR for each detector.

The GPR of two different dosimeters was assessed as a function of calibration field size. Since the plans delivered with each detector were distinct, direct comparison of the GPR for the same patient plan was not feasible. However, the response of the detectors could

https://doi.org/10.1017/S1460396923000444 Published online by Cambridge University Press

Table 1. Uncertainty budget

Uncertainty component	Relative uncertainty (%)	
	4DOCT	Arc Check
Machine O/P	0·21	0·2
Phantom setup	0·61	0·81
Plan delivery	0·1	0·68
Calibration factor	0·1	0·68
Total uncertainty k = 1, $u_c(y)$=	0·5262	1·491

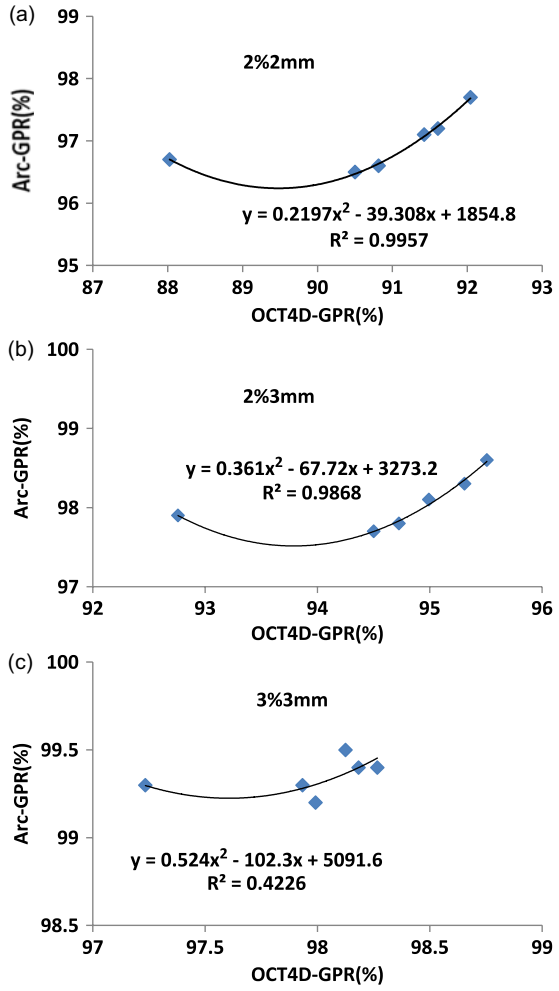

Figure 4. (a), (b), (c): Correlation of GPR between 4DOCT and Arc Check for (a = 2% 2mm, b = 2%3mm and c = 3%3mm)

be compared with respect to different CFs. Despite variations in the magnitude of GPR between the detectors, their responses were similar for different CFs; particularly at 2%2mm and 3%2mm gamma criteria.

The stability of GPR for the same plan was found to be sensitive to different calibration field sizes, especially at tighter gamma criteria (2%2mm and 3%2mm), as depicted in Figure 2 (4DOCT) and Figure 3 (Arc Check). When adopting a slightly relaxed criterion from 3%2mm to 3%3mm, changes in GPR for different calibration field sizes were less pronounced/sensitive, evident from the convergence of curves towards the 3%3mm gamma criteria in Figures 2 and 3.

In both detectors, the GPR exhibited an increasing trend for calibration field sizes less than 10 × 10 and a decreasing trend for sizes above 10 × 10. The overall conclusion drawn is that the GPR is better or more consistent when the dosimeter is calibrated with a field size less than 10 × 10 and greater than 4 × 4. These findings align with a study by Wei Luo et al. (2018)., which reported better GPR when the detector is calibrated with small field sizes (less than 10 × 10)[10]

Interestingly, a few cases deviated from these results, with 6 × 6 showing slightly higher GPR than 4 × 4 in 4DOCT. Similarly, in Arc Check, the GPR of 8 × 8 exhibited a slightly higher value compared to 4 × 4. To explain this, we hypothesised that any plan could achieve a better GPR if the detector is calibrated with a field size equal in dimension to the maximum number of repeated segments in the given plan. This aligns with a study in the literature, suggesting the use of effective field size rather than conventional field size for calibration and advising against a single common field-based calibration.[12]

Proof of hypothesis

Contrary to the reports by Wei Luo et al. (2018).[10] and Decabooter et al. (2021)12., our study aimed to address the previously unexplored question of why the gamma passing rate (GPR) is not stable concerning calibration field sizes. We developed a hypothesis and conducted a small proof of concept to shed light on this aspect.

A composite plan was generated in the treatment planning system (TPS), incorporating four beams. Among these, three beams shared the same dimensions, while the fourth had different dimensions. For example, the plan comprised three 4 × 4 fields and one 6 × 6 field. If the detector was calibrated with the 4 × 4 field, the measured dose closely matched the TPS. The results of the TPS and measured dose comparison by both detectors are presented in Tables 2 and 3. It is worth noting that all plan combinations favored our hypothesis, except for the first plan QA performed by Arc Check. The reason for this unfavourable result in the specific plan remains unknown. Nonetheless, these outcomes support our hypothesis, signalling the need for further investigation. To delve deeper into this phenomenon, future investigations could focus on extracting the dimensions of segments that are frequently repeated in a given clinical plan. Utilising the equivalent field of these segments as the calibration field may provide insights into the observed GPR variations.

As the GPR exhibits fluctuations with calibration field size, the reliability of PSQA results based on conventional calibration methods may be questionable. Despite achieving passing rates exceeding 90% through conventional calibration, we propose the adoption of a plan-specific calibration method for more accurate results. Plan-specific calibration field sizes have the potential to offer a better representation of clinical plans compared to conventional field sizes.

It is important to note that a major limitation of this study is that the proof of hypothesis has not been tested with real clinical plans. In future research, we aim to extend this study to validate our hypothesis using clinical plans, providing a more comprehensive understanding of the practical implications and potential improvements in PSQA procedures.

Need of plan specific calibration

The calibration coefficient of a chamber, determined under reference conditions, may not be directly applicable to the more complex fields encountered in IMRT/VMAT plans. When

https://doi.org/10.1017/S1460396923000444 Published online by Cambridge University Press

Table 2. Measured and TPS dose comparison as function of calibration field size in 4DOCT

Plan	TPS dose (Gy)	Measured dose (Gy		%diff	
A	1·05	(4 × 4) 1·054	(6 × 6) 1·06	−0·38 (4 × 4)	−0·95 (6 × 6)
B	1·067	(4 × 4) 1·072	(8 × 8) 1·08	−0·416 (4 × 4)	−1·22 (8 × 8)
C	1·082	(4 × 4) 1·087	(10 × 10) 1·1	−0·46 (4 × 4)	−1·66 (10 × 10

Plan A = Three 4 × 4 fields and one 6 × 6 field.
Plan B = Three 4 × 4 fields and one 8 × 8 field.
Plan C = Three 4 × 4 fields and one 10 × 10 field.
Field sizes in the bracket under the measured doses is the calibration field sizes.

Table 3. Measured and TPS dose comparison as function of calibration field size in Arc Check

Plan	TPS dose (Gy)	Measured dose (Gy)		%diff	
D	4·68	(4 × 4) 4·66	(10 × 10) 4·67	0·47 (4 × 4)	0·24 (10 × 10)
E	4·88	(4 × 4) 4·86	(10 × 10) 4·89	0·38 (4 × 4)	−0·25 (10 × 10)
F	4·93	(4 × 4) 4·89	(12 × 12) 4·91	0·76 (4 × 4)	0·45 (12 × 12)

Plan D = Three 4 × 4 fields and one 10 × 10 field.
Plan E = Three 10 × 10 fields and one 4 × 4 field.
Plan F = Three 12 × 12 fields and one 4 × 4 field.
Field sizes in the bracket under the measured doses is the calibration field sizes.

performing point dose measurements for PSQA, it is essential to consider plan-specific correction factors. Studies by Desai et al. (2019) have highlighted the importance of determining chamber correction factors tailored to specific plan types, such as step-and-shoot IMRT beams, VMAT beams, composite step-and-shoot plans and composite VMAT plans, using actual patient treatment data.[13] Their findings emphasise that a correction-free absorbed dose-to-water reading is not guaranteed even for large composite dose distributions, given the unique complexities of each plan.[13] A one-size-fits-all correction factor for all plans is not appropriate.

Alfonso et al. (2008) introduced a novel formalism[14] addressing nonstandard fields, categorising them into machine-specific reference (msr) for static small fields and plan-class-specific reference (pcsr) for composite fields (IMRT/VMAT plans). This formalism provides correction factors to mitigate differences between calibration and actual conditions in nonstandard fields. The magnitudes of these factors vary with field sizes, cases and dosimeters used. As highlighted by Fitriandini et al. (2019), these factors should be considered as additional correction factors for PSQA.[15]Decabooter et al. (2021)[12] investigated the GPR for stereotactic radiosurgery (SRS) cases using the PTW 1600 SRS detector. They deviated from the manufacturer's recommended calibration method, which employs simple square fields. Instead, they opted for an average equivalent square field size for detector calibration, reporting that conventional field-based calibration could lead to systematic dose deviations of up to 4·1%.

From the above discussion, it is very clear that every plan required its own correction factor. In our study, we presented that the improper calibration field size also causes error in PSQA, and the results aligns with literatures.[10,12,13,15] However, it is essential to note that our study specifically addresses the effect of calibration field sizes on PSQA, and further research may be required to comprehensively explore other aspects of the PSQA process.

Conclusion

This study investigated the influence of calibration field size on the GPR for PSQA using two different detectors. Both detectors exhibited variations in GPR across a range of calibration field sizes. Given the irregular and complex nature of IMRT fields, it was inferred that a plan-specific calibration method might be necessary.

The proposed approach for selecting calibration field sizes involves three steps: (1) identifying the size of the most repeated segments in the plan, (2) determining the equivalent square field size of these repeated segments and (3) utilising this field size for calibrating the QA device. While it is acknowledged that this suggested calibration method may pose challenges in regular clinical workflows, the study identified the underlying reasons for GPR instability with different calibration field sizes. These points to the need for further research to explore practical ways of implementing this calibration method.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1460396923000444.

Acknowledgements. The authors express their gratitude to Narmada C, Senior Physicist & Application Specialist at PTW-Dosimetry India Pvt Ltd, for her invaluable contributions to editing the manuscript. The authors also extend their thanks to CIHSR Hospital, Dimapur, Nagaland, India, and Meherbai Tata Memorial Hospital, Dhatkidih, Jamshedpur, Jharkhand for providing the radiotherapy facility essential for data collection.

Competing interests. There are no conflicts of interest.

References

1. Dogan N, Mijnheer BJ, Padgett K, et al. AAPM task group report 307: use of EPIDs for patient-specific IMRT and VMAT QA. Med Phys 2023; 50 (8): e865–903.
2. 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.
3. U'wais FA, Radzi Y, Rizan NN, et al. Validation of a digital method for patient-specific verification of VMAT treatment using a 2D ionisation detector array. Radiat Phys Chem 2023; 202: 110536.

4. Infusino E, Ianiro A, Luppino S, et al. Evaluation of a planar diode matrix for SRS patient-specific QA in comparison with GAFchromic films. J Appl Clin Med Physics 2023; 24 (8): e13947.
5. Ali AM, Greenwood JB, Varasteh M, et al. Analysis of the interplay effect in lung stereotactic ablative radiation therapy based on both breathing motion and plan characteristics. J Radiother Pract 2023; 22 (e75): 1–11
6. Kunii Y, Tanabe Y, Higashi A, et al. Effects of high-resolution measurements between different multi-row detectors on volumetric modulated arc therapy patient-specific quality assurance. Int J Radiat Research 2023; 21 (3): 413–419.
7. James S, Al-Basheer A, Elder E, et al. Evaluation of commercial devices for patient specific QA of stereotactic radiotherapy plans. J Appl Clin Med Physics 2023; 9: e14009.
8. Low DA, Harms WB, Mutic S, et al. A technique for the quantitative evaluation of dose distributions. Med Phys 1998; 25 (5): 656–661.
9. Hussein M, Adams EJ, Jordan TJ, et al. A critical evaluation of the PTW 2D-ARRAY seven29 and OCTAVIUS II phantom for IMRT and VMAT verification. J Appl Clin Med Phys 2013; 14 (6): 4460.
10. Luo W, Meng Y, Westlund SB Dose calibration uncertainty and plan-specific dose calibration for IMRT QA. Biomed Phys Eng Express 2018; 5: 015024. doi: 10.1088/2057-1976/aae410
11. Chaswal V, Weldon M, Gupta N, et al. Commissioning and comprehensive evaluation of the ArcCHECK cylindrical diode array for VMAT pretreatment delivery QA. J Appl Clin Med Phys 2014; 15 (4): 4832.
12. Decabooter E, Swinnen AC, Öllers MC, et al. Operation and calibration of the novel PTW 1600SRS detector for the verification of single isocenter stereotactic radiosurgery treatments of multiple small brain metastases. Br J Radiol 2021; 94 (1123): 20210473.
13. Desai VK, Labby ZE, Hyun MA, et al. VMAT and IMRT plan-specific correction factors for linac-based ionization chamber dosimetry. Med Phys 2019; 46 (2): 913–924.
14. Alfonso R, Andreo P, Capote R, et al. A new formalism for reference dosimetry of small and nonstandard fields. Med Phys 2008; 35 (11): 5179–5186.
15. Fitriandini A, Wibowo WE, Pawiro SA Determination of composite field correction factor kpcsr,msr for Tomotherapy. J Phys: Conf Ser 2019; 1248: 012058.

https://doi.org/10.1017/S1460396923000444 Published online by Cambridge University Press