Evaluating suggested stricter gamma criteria for linac-based patient-specific delivery QA in the conventional and SBRT environments


Fox Chase Cancer Center, Department of Radiation Oncology, 333 Cottman Avenue, Philadelphia, PA 19111, United States

ARTICLE INFO

Keywords:
Gamma criteria
IMRT
VMAT QA

ABSTRACT

Purpose: To evaluate AAPM TG-218 recommended tolerances for IMRT QA for conventional and SBRT delivery.

Methods: QA analysis was repeated for 150 IMRT/VMAT patients with varying gamma criteria. True composite delivery was utilized, corrected for detector and output variation. Universal tolerance (TL\text{univ}) and action limits (AL\text{univ}) were compared with statistical process control (SPC) TL\text{SPC} and AL\text{SPC} values. Analysis was repeated as a function of plan complexity for 250 non-stereotactic body radiotherapy (SBRT) VMAT patients at 3%/2mm and a threshold of 10% and for 75 SBRT VMAT patients at 2%/2 mm and a threshold of 50% with results plotted as a function of PTV volume. Regions of failure were dose-scaled on the planning CT data sets based on delivery results.

Results: The IMRT/VMAT TL\text{SPC} and AL\text{SPC} for gamma criteria of 3%/3 mm were 96.5% and 95.6% and for 3%/2 mm were 91.2% and 89.2%, respectively. Correlation with plan complexity for conventional fractionation VMAT was "low" for all sites with pelvis having the highest r value at ~0.35. The equivalent SBRT PTV diameter ranged from 2.0 cm to 5.6 cm. Negative low correlation was found for 38 of 75 VMAT cases below AL\text{univ}.

Conclusions: The AL\text{univ} and AL\text{SPC} are similar for 3%/2 mm. However, our 5% failure rate for AL\text{univ}, may result in treatment start delays approximately 2 times/month, given 40 new cases/month. VMAT QA failure at stricter criteria did not correlate strongly with plan complexity. Site-specific action limits vary less than 3% from the average. SBRT QA results do not strongly correlate with target size over the range studied.

Introduction

The AAPM TG-218 report provides excellent guidance on the procedures to be followed when performing patient-specific IMRT delivery QA [1]. In addition, the use of statistical process control (SPC) analysis allows for evaluation of the “strength” of a QA program as compared to universally accepted tolerance (TL\text{univ}) and action limits (AL\text{univ}). It has been recommended that the gamma criteria including percent dose (%) and distance to agreement (DTA), to be evaluated at the aforementioned limits, be lowered from the commonly used 3%/3mm to 3%/2mm (both at a dose threshold of 10%). The application of this recommended stricter gamma criteria (ie. lower DTA) has practical ramifications that should be understood prior to implementation.

The use of the aforementioned TL\text{univ} and AL\text{univ} concepts assumes that these limits are coincident with the goals of the treatment plan being evaluated. For scenarios where this may not be the case such as SBRT, the report suggests that tighter tolerances be considered. In addition, stricter gamma criteria should be used to detect subtle regional errors. SBRT planning typically involves smaller targets and results in less homogeneous dose distributions than more conventionally fractionated IMRT/VMAT plans. An understanding of the clinical significance of this potential decrease in passing rate is essential for SBRT treatment delivery.

The AAPM TG-218 report indicates that a source of variation in IMRT QA measurement can be attributed to variation in plan complexity, specifically between different body sites. It is suggested that it is good practice to determine tolerance limits for cases of high vs. low complexity (ie. head and neck and prostate IMRT cases). To this end it is useful to have an understanding of the effects of plan complexity on delivery QA gamma passing rates.

The purpose of this work is to evaluate, based on the current strength of our QA process, the AAPM TG-218 recommendations for QA
tolerances for patient-specific intensity modulated treatments specifically with regard to gamma analysis. We thoroughly evaluate whether or not stricter gamma criteria may result in delays in patient treatment initiation. We evaluate whether or not the use of the stricter gamma criteria for the SBRT setting with the combination of smaller targets and non-homogeneous dose distributions as well as finite detector spacing of 2D arrays may contribute to potential lower passing rates. Additionally, we evaluate whether or not VMAT plan complexity has an effect on gamma passing rates and whether or not site-specific tolerance limits are necessary. Additionally, this work will help illustrate the clinical significance of deviations in delivery detected through the use of stricter gamma criteria for IMRT/VMAT in the conventional and SBRT environments.

Methods and materials

IMRT/VMAT: Conventional fractionation

Patient-specific ion chamber array-based delivery QA analysis was repeated with varying gamma criteria for 150 IMRT/VMAT patients previously treated between May 2018 and January 2019. For all cases throughout this study our planning techniques varied minimally between multiple planners and physicians and departmental site-specific dose-volume plan acceptance criteria remained consistent over the given time periods. All delivery QA measurements were obtained using one of two IBA MatriXX Evolution and MultiCube Lite phantom combinations and analyzed with the OmniProTM GammaRT software (IBA Dosimetry America, Bartlett, TN). Both systems were calibrated according to the manufacturer’s instructions prior to use. This 2D array utilizes 1020 ionchambers with 0.76 cm center-to-center spacing. All measurements were interpolated to 1 mm resolution prior to analysis. True composite delivery was utilized, with all analysis performed in absolute dose and corrected for detector and output variation with angular correction applied to all IMRT cases [2]. Angular correction was based on measurements made on our device using a 10 cm × 10 cm field size at varying beam angles and applied using in-house developed software that corrects for machine output variation as well. TG-218 indicates that “…angular dependence may be smeared out when using beams from many angles, such as with VMAT delivery…” [1]. It has been our practice not to apply angular correction to VMAT measurements for gamma analysis and our results support this. The software used does not automatically optimize the position of the dose distribution to minimize spatial uncertainty between measured and planned dose distributions. Any adjustment in the position of the measured dose distribution was made manually in the x and y direction on the coronal plane and limited to perceived phantom setup error (typically 1 mm in either direction). Passing rates for Tnuniv and Aluniv (≥95% and 90%, respectively) were determined following global normalization and a low dose threshold of 10%. Center-specific SPC TL (TLSpc) and AL (ALspc) values were determined for comparison with the Tnuniv and A1univ. The center-specific values were determined according to the method given in the AAPM-TG 218 report using the following equation,

\[ \Delta A = \beta \sqrt{\sigma^2 + (\bar{x} - T)^2} \]  

where \( \Delta A \) is the difference between upper and lower action limits and is typically written as ± A/2. T is the process target value and \( \sigma^2 \) and \( \bar{x} \) are the process variance and mean, respectively. As our target value for IMRT QA gamma passing rate was known (ie. 100%), this value was used for T. \( \beta \) is a constant combining the cutoff for an acceptably performing process and a factor that balances type I errors (rejecting the null hypothesis when it is true). For IMRT QA the null hypothesis is that the process is unchanging. As suggested based on current information, a value of 6.0 was used. Control chart limits determined from an I-chart of individual QA measurements were used as tolerance limits. The I-chart is a statistical tool used to identify QA measurements displaying abnormal process behavior. The I-chart has upper and lower limits or control limits and a center line, all calculated from the IMRT QA measurement data. Abnormal behavior is indicated when an IMRT QA measurement results in a value outside the upper or lower limits on the I-chart. The center line and upper and lower limits are given as follows:

\[ \text{center line} = \frac{1}{n} \sum_{i=1}^{n} x_i \]  

\[ \text{upper control limit} = \text{center line} + 2.660 \cdot \frac{mR}{\sqrt{n}} \]  

\[ \text{lower control limit} = \text{center line} - 2.660 \frac{mR}{\sqrt{n}} \]  

where x is an individual QA measurement and n is the total number of measurements and mR is the moving range.

To evaluate the effects of increasing plan complexity on passing rate an approximation modulation scaling factor (MSFapprox) was used as a surrogate, defined as:

\[ \text{MSFapprox} = \frac{\text{MU}_{\text{IMRT}}}{\text{Dose}_{\text{frac}}(\text{approx})} \]  

where MSF_{IMRT} is the required MU for a single fraction of the IMRT/VMAT plan and Dose_{frac(approx)} is the associated prescribed dose for the same fraction. This approximation is based on the general idea that complex plans are derived from complex fluence distributions requiring leaf sequences containing many small, high-MU segments to deliver. Passing rate as a function of MSFapprox was evaluated for both IMRT and VMAT delivery and gamma criteria of 3%/3mm and 3%/2mm using a 10% threshold.

The average time between QA delivery and patient start was also determined for these patients. This time was used to calculate the frequency of patient start delays assuming the AL defines a “no treatment” threshold.

Conventional fractionation VMAT and site-specific tolerance limits

TG-218 suggests that plan complexity, and subsequently QA delivery passing rates, vary by treatment site and that it is good practice to calculate site-specific tolerance limits. To evaluate this claim, a series of 250 VMAT cases treated between May 2014 and February 2021, previously evaluated at 3%/3mm and a 10% threshold were re-evaluated at 3%/2mm and a 10% threshold. VMAT delivery was chosen as it was found to have a higher ALspc than IMRT delivery in our QA system. Based on our departmental delivery QA experience we hypothesized that site-specific tolerance limits are unnecessary. Cases used for evaluation were selected as follows: H&N (n = 76), Thorax (n = 40), Abdomen (n = 61) and Pelvis (n = 73). Each treatment site was evaluated as a function of MSFapprox as defined above. Statistical Process Control (SPC) methods were used to compare the action limit of each subset to the universal action limit (Aluniv) of 90%.

VMAT SBRT

TG-218 suggests gamma criteria for SBRT be tighter or stricter than those used for non-SBRT delivery. To this end we selected 75 VMAT SBRT cases treated between October 2017 and December 2019 having previously passed routine criteria of 3%/3mm at 10% threshold and reanalyzed at 2%/2mm and a threshold of 50%. All cases were evaluated as a function of PTV volume. Our reasoning for this threshold was based on the rapid dose falloff exploited by SBRT planning. With all evaluations globally normalized to a value within 10% of the maximum dose in the measured data set, resulting values below 50% would be less likely
to be clinically significant (i.e. less likely to exceed normal tissue dose volume limits). In addition, by decreasing the total number of pixels evaluated in the low dose regions, we can decrease the likelihood of a delivery achieving an acceptable gamma passing rate while failing in clinically significant regions, (e.g. the target). Acknowledging the difficulty in representing the multitude of complex PTV shapes encountered, the equivalent diameter was determined by converting the PTV volume to a sphere for illustration purposes. The equivalent diameter ranged from 2 to 5.6 cm corresponding to CTV/ITV diameters from 1 to 4.6 cm using a 0.5 cm uniform PTV margin, for example. Due to potential difficulties measuring dose distributions for smaller targets due to the detector spacing of our device, data for PTVs < 2 cm diameter are not presented in this work. Given the 0.76 cm center-to-center spacing of our detector array, our hypothesis was that we would see decreasing gamma passing rates as a function of decreasing PTV diameter. The upper limit of 5.6 cm diameter corresponds approximately with the largest targets selected for SBRT at our institution.

In order to determine the clinical relevancy of dose levels of points failing the ALuniv as suggested in TG-218, the cases were evaluated for failures inside and outside the PTV. Delivery QA and gamma analysis were performed with failing regions identified in the axial, sagittal and coronal planes of the original treatment planning CT using isocenter as reference. The corresponding dose was adjusted (scaled) based on the measured vs. planned phantom dose ratio at the same coordinates.

Results

**IMRT/VMAT: Conventional fractionation**

From the 150 IMRT/VMAT cases evaluated, the percentage of pixels falling between TLuniv and ALuniv as well as below ALuniv are presented in Table 1. Additionally, the center-specific TLSCP and ALSCP are determined as demonstrated in TG-218 using the increased number of patients and compared with the universal limits. The average global normalization was 95.2% of the maximum dose in the measurement plane. The average passing rates for the total cohort for gamma criteria of 3%/3mm and 3%/2mm were 98.8% and 96.5%, respectively. The percentage of cases failing TLuniv for gamma criteria of 3%/3mm and 3%/2mm were 1% and 1% and failing ALuniv were 0% and 5%, respectively. During the evaluation of the 5% of cases failing the ALuniv at 3%/2mm, it was found that 7 patients were associated with IMRT delivery and 1 patient received VMAT delivery. Consultation with the attending physicians with respect to the OAR or target involved, indicated that none of the patients would have had their starts delayed based on these findings alone. The center-specific TLSCP and ALSCP for gamma criteria of 3%/3mm and 3%/2mm were 98.6%, 94.9% and 95.8%, 86.1%, respectively. While the ALSCP of 86.1% for 3%/2mm is reasonably close to the ALuniv of 90% it is indicative of the need for improvement in our overall delivery QA process. This improvement may include changing our routine phantom alignment methodology to be image-based rather than using positional lasers and linac crosshairs. Changing to conebeam computed tomography (CBCT) would be more in line with how our patients are aligned on a daily basis. Optimal plane selection to position the detector array within the high dose low gradient region for routine delivery QA may also help improve the process. Improvement in our angular correction methodology for IMRT delivery to include potential change as a function of segment size as well as where the segment(s) is located within the array may also be beneficial. The latter 2 topics are the subjects of separate ongoing studies at our institution. The average of the maximum gamma values found during analysis for 3%/3mm and 3%/2mm were 1.46 and 1.73, corresponding to 1.38%/1.38 mm and 2.19%/1.46 mm beyond limits, respectively. The average percentage of pixels exceeding a gamma value of 1.5 was 0.04% and 0.27%, respectively. Evaluating the IMRT and VMAT cases separately yields ALSCP values of approximately 82.2% and 91.6%, respectively, as depicted in Fig. 1. The lower value for IMRT indicates that VMAT QA performs better than IMRT with respect to SPC in our process. The percentage of pixels failing below ALuniv as well as ALSCP as a function of decreasing gamma criteria are presented in Table 2.

The percentage of pixels passing gamma criteria of 3%/3mm and 3%/2mm as a function of MSFapprox for IMRT only, are given in Fig. 2. The Pearson product moment r value for 3%, 3 mm was −0.35 (p = 0.002) indicating “low” correlation. For the purposes of this study the following categories were used to describe the strength of correlation: 0.00–0.25 (little if any), 0.26–0.49 (low), 0.50–0.69 (moderate), 0.70–0.89 (high) and 0.90–1.00 (very high). The r value for 3%/2mm is −0.25 (p = 0.031) indicating “little if any” correlation. The percentage of pixels passing gamma criteria of 3%/3mm and 3%/2mm as a function of MSFapprox for VMAT only, are given in Fig. 3. The r values for 3%/3mm and 3%/2mm were −0.16 (p = 0.170) and −0.19 (p = 0.103), respectively, indicating “little if any” correlation.

**Conventional fractionation VMAT and site-specific tolerance limits**

The correlation of gamma passing rates at 3%/2mm and a 10% threshold, as a function of MSFapprox, demonstrated in Fig. 4, was “low” for all sites with pelvis having the highest r value at −0.35 (p = 0.002). All metrics evaluated are illustrated in Table 3. The site-specific action limits differ from the ALuniv by 0.9, −1.1, −2.3, and −4.4% for the Abdomen, H&N, Pelvis and Thorax, respectively. The variance in MSFapprox values between groups reached a probability level of less than 1% indicating that plan complexity is different between groups. However, the passing rates did not reach significance indicating that this parameter does not vary with plan complexity.

**VMAT SBRT**

Fig. 5 illustrates the initial passing rates at 3%/3mm (diamonds), the overall results for the stricter gamma criteria of 2%/2mm (hollow squares) and the values failing the ALuniv (90%) at the stricter criteria (triangles). Approximately 26% of the cases were below the ALuniv with increasing failure with increasing equivalent diameter. The correlation between passing rate and target size was low (r = −0.31, p = 0.006).

Fig. 6 illustrates the results for dose value comparisons failing criteria within the PTV only. Fifty-seven percent of these cases were evaluated at 2 regions failing criteria when compared to calculation. Assuming the ablative intent of SBRT and resultant heterogenous dose distributions within the target, it is found that the comparison of failed regions with the maximum percentage of prescribed dose in the approved treatment plans would not warrant treatment cancellation or delay. This process was repeated for regions outside the PTV with results illustrated in Fig. 7. The 50% threshold was chosen to include critical structures adjacent to the target. Four of the 7 regions fell below approved doses (2 cases had 2 regions). Three of 7 cases resulted in regions exceeding dose as depicted in the TPS (max deviation 20.2%); however, none of the regions resulted in doses that would exceed OAR tolerance.

**Table 1**

The percentage of pixels falling between TLuniv and ALuniv (% below TLuniv) as well as those below ALuniv, TLSCP and ALSCP values as a function of decreasing gamma criteria are also presented.

<table>
<thead>
<tr>
<th>IMRT/VMAT</th>
<th>3%/3mm (%)</th>
<th>3%/2mm (%)</th>
<th>2%/2mm (%)</th>
<th>2%/1mm (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(% below TLuniv)</td>
<td>1</td>
<td>19</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>(% below ALuniv)</td>
<td>0</td>
<td>5</td>
<td>36</td>
<td>94</td>
</tr>
<tr>
<td>TLSCP (%)</td>
<td>98.6</td>
<td>95.8</td>
<td>89.9</td>
<td>75.4</td>
</tr>
<tr>
<td>ALSCP (%)</td>
<td>94.9</td>
<td>86.1</td>
<td>68.7</td>
<td>27.9</td>
</tr>
</tbody>
</table>

Combined IMRT/VMAT TL and AL for stricter gamma criteria.
Discussion

IMRT/VMAT: Conventional fractionation

Throughout this work we have chosen to concentrate mainly on the AL boundary vs. the TL. We consider the AL a “no-treatment lower limit” for passing rate indicating that our process may be out of control. The information obtained from the gamma analysis in one or two irradiation planes (requiring table height adjustment between planes and phantom re-irradiation) is insufficient to determine the clinical significance of the QA failures. Additionally, it has long been the practice at our center to investigate cases further on an individual basis if there is variance from the expected value for a given class of cases (i.e. H&N, prostate, etc.), regardless if TL has been met. We have also found value in the SPC

<table>
<thead>
<tr>
<th>(%) below AL&lt;sub&gt;univ&lt;/sub&gt;</th>
<th>IMRT only</th>
<th>3%/3mm</th>
<th>3%/2mm</th>
<th>2%/2mm</th>
<th>2%/1mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL&lt;sub&gt;SPC&lt;/sub&gt; (%)</td>
<td>IMRT only</td>
<td>93.8</td>
<td>82.2</td>
<td>61.7</td>
<td>13.4</td>
</tr>
<tr>
<td>VMAT only</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>AL&lt;sub&gt;SPC&lt;/sub&gt; (%)</td>
<td>VMAT only</td>
<td>96.1</td>
<td>91.6</td>
<td>77.8</td>
<td>46.0</td>
</tr>
</tbody>
</table>

IMRT and VMAT specific AL for stricter gamma criteria.

Table 2

Fig. 1. Percentage of pixels passing TL<sub>univ</sub> and AL<sub>univ</sub> as a function of gamma criteria of increasing strictness. Evaluation separated into 75 IMRT and 75 VMAT cases using a low dose threshold of 10% throughout.

Fig. 2. Percentage of pixels passing gamma criteria as a function of IMRT plan complexity. The mean MSF<sub>approx.</sub> values are for illustrative purposes only.
method in determining the strength of our process by comparing our AL_{SPC} to the AL_{univ}. Our combined IMRT/VMAT AL_{SPC} was 94.9% and 86.1% for 3%/3mm (10% threshold) and 3%/2 mm, respectively. While 86.1% is “close to” the 90% AL_{univ} value proposed in TG-218, it indicates improvement may be needed. This improvement may include image-based phantom alignment, optimization of phantom measurement

Table 3
Metrics for conventional fractionation VMAT, site-specific delivery QA. CV = coefficient of variation, r = Pearson product moment correlation.

<table>
<thead>
<tr>
<th>Site</th>
<th>TL_{SPC}</th>
<th>AL_{SPC}</th>
<th>Abs diff (AL_{univ}-AL_{SPC})</th>
<th>CV</th>
<th>r</th>
<th>p</th>
<th>Average MSF_{approx}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdomen</td>
<td>92.8</td>
<td>90.9</td>
<td>0.90</td>
<td>2.12</td>
<td>-0.30</td>
<td>0.020</td>
<td>2.8</td>
</tr>
<tr>
<td>H&amp;N</td>
<td>91.0</td>
<td>88.9</td>
<td>1.11</td>
<td>2.18</td>
<td>-0.30</td>
<td>0.008</td>
<td>2.6</td>
</tr>
<tr>
<td>Pelvis</td>
<td>89.3</td>
<td>87.7</td>
<td>2.32</td>
<td>2.63</td>
<td>-0.35</td>
<td>0.002</td>
<td>3.4</td>
</tr>
<tr>
<td>Thorax</td>
<td>88.7</td>
<td>85.6</td>
<td>4.40</td>
<td>3.08</td>
<td>-0.28</td>
<td>0.080</td>
<td>3.2</td>
</tr>
<tr>
<td>TOTAL</td>
<td>90.1</td>
<td>88.4</td>
<td>1.60</td>
<td>2.50</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Treatment site-specific TL and AL (VMAT only).
plane selection as well as a more robust angular correction method. Separating this value by delivery method yields 82.2% and 91.6% AL for IMRT and VMAT at 3%/2mm, respectively. The lower value for IMRT is thought to be related to our implementation of device angular correction and its dependency on segment size and position. This dependency is being investigated further.

From Fig. 1 it can be seen that while adopting the AAPM TG-218 suggested gamma criteria of 3%/2mm (10% threshold) is possible, further reductions would require significant improvement in our QA process. As it stands, the suggested criteria are associated with practical ramifications. As previously mentioned we consider the AL to be a no-treatment boundary. From Table 1 it is seen that our QA process results in values failing this boundary approximately 5% of the time for a gamma of 3%/2mm. In addition, we have found that plan approval and subsequent delivery QA are performed the evening prior to treatment start approximately 40% of the time. If we assume 1700 patients are treated annually with 70% via IMRT/VMAT delivery, this would indicate approximately 2 patients/month having their treatment start date delayed. This will inevitably have ramifications for the patient ranging from scheduling issues to coordination with chemotherapy delivery. Failure of the AL for the current criteria of 3%/3mm is found to be 0%. Should the decrease in gamma criteria be beneficial to the patient these ramifications would of course, be warranted.

We found no meaningful correlation for this patient population supporting the idea, as reported by others, that delivery failure as depicted through gamma analysis does not correlate well with clinically meaningful endpoints even for stricter criteria [5–11]. In fact, both false negative and false positive results are possible, rendering plan deliverability acceptance to the evaluation of a passing rate, questionable [9]. This should not reflect negatively on the usefulness of the gamma index method as a tool but may in part be a function of how the QA system is commonly implemented [12] (i.e. measurements made in the high dose low gradient region) where the measurement apparatus may or may not be coincident with the plane containing organs at risk. Given the 2D-nature of available detector arrays, it is impractical to acquire enough information to potentially lead to markedly improved significance.
It has also been suggested that failure rates may increase as a function of plan complexity [1]. Figs. 2 and 3 illustrate failure rate as a function of MSF\textsubscript{approx} for IMRT and VMAT delivery, respectively. Over the ranges studied including up to 8 for IMRT and 5 for VMAT, this was not shown to be the case for gamma criteria of 3%/3mm or 3%/2mm with “low” correlation being the highest level attained. The reduction in gamma criteria does not appear to improve this correlation appreciably given the manner in which we have described plan complexity.

We have found minimal correlation between gamma analysis and clinical endpoints, as well as the lack of correlation with plan complexity. Given our clinical practice as described, we believe the suggested stricter gamma criteria to be unwarranted. These findings may not be translatable for other scenarios. Knowing this, without correlation between gamma analysis results and patient outcomes or at least a 3D-reconstructed plan deliverability-based rejection process, arbitrarily decreasing gamma criteria for plan acceptance may do nothing more than add unnecessary treatment start delays and additional, unfounded work for the medical physicist/dosimetrist. Perhaps patient-specific delivery QA and subsequent gamma analysis, following establishment of an appropriate QA process through the SPC method, can be used to verify endpoints that may be coarse but crucial such as:

- Is there evidence of modulation (leaf sequencing)?
- Has the expected absolute dose been delivered?
- Does the shape of the delivered dose distribution resemble the planned distribution?
- Are expected bulges and concavities in the distribution present in the appropriate locations?
- Do the measured and TPS-calculated isodoses coincide appropriately over the range evaluated?

This is fundamentally based on the delivery of physical dose and issues will not resolve if measurements are only made once at limited points. The question is whether gamma analysis is simply to help avoid RT accidents (severe dose errors or target misses) or can potentially improve clinical outcome. For example, a point dose measurement with a 5% dose action level could potentially avoid significant dose errors, but reducing it to 2% would not improve the clinical outcome.

**Conventional fractionation VMAT and site-specific tolerance limits**

The TG-218 report suggests that plan complexity varies as a function of treatment site and infers good practice to include different tolerance limits for H\&N vs. prostate cases, as an example. We evaluated gamma passing rates at 3%/2mm (10% threshold) for VMAT delivery of H\&N, Pelvis, Thorax, and Abdomen plans, all with respect to plan complexity as previously defined. We found correlation with plan complexity to be low for all QA results across all groups. In fact, H\&N demonstrated the lowest plan complexity with an average MSF\textsubscript{approx} of 2.6, and the second highest AL\textsubscript{SPC} with a value of 88.9% with only abdomen having a higher value. AL decreases from 90.9, 88.9, 87.7 and 85.6% for Abdomen, H\&N, Pelvis and Thorax, respectively. However, plan complexity as measured by the MSF\textsubscript{approx} is actually greater for Pelvis (3.4) than Thorax (3.2). Given the critical structure limits, particularly low-dose limits for lung when treating the thorax, and the increased use of partial arcs, the changes in SPC derived AL may be device related and not necessarily plan complexity related as defined here.

VMAT is rarely used for thoracic treatment outside the SBRT setting at our institution. This is evidenced by the lower number of cases available for evaluation in this body site as well as the higher coefficient of variation (CV) of 3.08% where.

\[
CV = \left(\frac{\text{standard deviation}}{\text{mean}}\right) \times 100.
\]

Removing thorax from this study would result in a TOTAL CV of approximately 2.31 and tighter variation between subgroups, further supporting a singular TL and AL for all groups. As it stands the Thorax delivery QA values are more variable relative to their mean than Pelvis, Abdomen or H\&N.

**VMAT SBRT**

The low correlation of passing rate with PTV volume, at the reduced gamma criteria, was not clinically meaningful. While opposite our initial hypothesis, this may be attributed to more complex shapes with larger volume targets. It seems intuitive that larger volume targets may share proximity with critical structures with strict dose volume limits and may result in concave dose distributions. The high dose gradients associated with these concavities and resulting from the conformity of the dose distribution, may present more difficulties for arrays with lower detector resolution. Additionally, given the 0.76 cm center-to-center spacing in our chamber array, the higher passing low-volume targets, typically more spherical or ellipsoidal in shape, may reflect positively with respect to the interpolation method used with this device. It is of note that smaller diameter targets present challenges for delivery QA and may in part be a function of detector resolution as reported by others [13,14]. Raw readings are measured at the detector resolution defined by the inherent detector spacing. The treatment planning system exported dose distribution is at a resolution of 1 mm × 1 mm. If a “peak
in the dose distribution occurs within the boundaries of the chambers (between detectors), the planning system will export this; however, the peak will not be picked up by the chambers during measurement. All smoothing of the measurement to finer resolution is based on interpolation. If none of the chambers detect the high dose region(s) then the interpolated isodose lines cannot contain it. Any following gamma comparison will then fail (depending on the magnitude of the dose difference between the exported peak and the interpolated dose value at the same position). One practical way of solving this is to shift the array to assure the dose peak coincides with a chamber location. This value is then available in the interpolations and helps to minimize this effect.

The lack of correlation of gamma results with clinical endpoints remains an issue even with SBRT [15–18]. Xia et al. evaluated 4 different QA systems for SRS and SBRT delivery and found that it is possible to apply stricter, device-specific gamma criteria for analysis with acceptable action and tolerance limits. They demonstrated reducing the criteria from the commonly used standard of 3%/3mm to a range down to 3%/1mm; however, they found no correlation with plan complexity [18]. In the current work we have also demonstrated reducing the gamma criteria to 2%/2mm at a 50% threshold with acceptable action limits with no correlation with plan complexity. In addition, we have found no correlation with PTV volume nor would cases failing these reduced criteria have resulted in treatment delays, further illuminating the lack of correlation with clinical endpoints.

There has been work reported that would allow for a migration to delivery-based QA resulting in 3D dose calculated for comparison with the initial, physician-approved treatment plan through DVH analysis [9,19–27]. This would allow the physician to make decisions based on dose-volume limits and dose distributions, similar to initial plan acceptance, providing a direct comparison with clinical endpoints without having to decipher potential deviations of dose or DTA in the extremes for low and high gradient regions. Given the work presented here we would advocate that the excellent methods described in the TG-218 report be followed to achieve a level of conformity between centers and that the SPC methods be utilized to assure that the strength of individual programs compare favorably with the universal action limit definition as described. However, we fail to find advantage in decreasing the gamma criteria from 3%/3mm to 3%/2mm (threshold = 10%).

As indicated previously, our delivery QA process follows the recommendations of the AAPM TG-218 report. In addition, the software/hardware system used is associated with an answer of “Yes” to all questions in the vendor survey section of that report except for providing auto registration or assuming the center of each image as a common point. To this end we believe this work to be generalizable for centers practicing similarly and with comparable QA systems but would suggest caution and evaluation of a center-specific patient cohort to verify.

Conclusions

From this study it appears possible to tighten our gamma criteria from 3%/3mm to 3%/2mm (10% threshold) as suggested in AAPM TG-218. However, the change from 0% to 5% of cases failing the A\(_{\text{univ}}\) could result in treatment starts being delayed approximately 2 times/month, given a monthly average of 40 IMRT/VMAT new patients where delivery QA was performed the night prior to treatment initiation. In Addition, none of the regions failing the gamma criteria in these cases correlate with clinically significant endpoints. Our findings from SPC limits derived from our clinical data sets demonstrate our fixed-beam IMRT delivery process to be inferior to that of VMAT as illustrated through the greater deviation from the A\(_{\text{univ}}\) and warrants further investigation. In this study we found that failure to pass gamma criteria as a function of plan complexity, as described by the MSF\(_{\text{approx}}\), exhibited little to no correlation for IMRT or VMAT delivery over the ranges studied. Additionally, our analysis of failure rates as a function of VMAT plan complexity across different body sites did not support site-specific action limits.

Following VMAT SBRT delivery QA evaluation using the array specified, it was found that results do not strongly correlate with target size over the range studied and it can be argued that no case in this study would have been delayed for gamma failure either inside or outside the PTV. Variance resulting in higher than predicted dose outside the target was found to be acceptable in all cases.

We find the methodology described in TG-218 to be an excellent guide to assure more uniform IMRT/VMAT delivery QA results throughout the radiation oncology community. However, through this work we demonstrate that the suggested stricter gamma criteria may not strongly predict clinically meaningful delivery errors.

Author contribution statement

All authors contributed to the acquisition, analysis and interpretation of the data associated with this work and have critically revised drafts and given approval for the final version for submission for publication.

The authors have no competing interests to declare.

This publication was supported by grant number P30 CA06927 from the National Cancer Institute, NIH. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References


