Australas Phys Eng Sci Med
DOI 10.1007/s13246-014-0274-9

SCIENTIFIC PAPER

Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results

S. B. Crowe · T. Kairn · J. Kenny · R. T. Knight · B. Hill · C. M. Langton · J. V. Trapp

Received: 25 January 2014 / Accepted: 22 April 2014

Abstract The planning of IMRT treatments requires a compromise between dose conformity (complexity) and deliverability. This study investigates established and novel treatment complexity metrics for 122 IMRT beams from prostate treatment plans. The Treatment and Dose Assessor software was used to extract the necessary data from exported treatment plan files and calculate the metrics. For most of the metrics, there was strong overlap between the calculated values for plans that passed and failed their quality assurance (QA) tests. However, statistically significant variation between plans that passed and failed QA measurements was found for the established modulation index and for a novel metric describing the proportion of small apertures in each beam. The 'small aperture score' provided threshold values which successfully distinguished deliverable treatment plans from plans that did not pass QA, with a low false negative rate.

Keywords Quality assurance · Beam complexity · Radiation therapy

This study was supported by the Australian Research Council, the Wesley Research Institute, Premion and the Queensland University of Technology (QUT), through linkage grant number LP110100401.

S. B. Crowe (✉) · T. Kairn · C. M. Langton · J. V. Trapp
Science and Engineering Faculty, Queensland University of Technology, Brisbane, QLD, Australia
e-mail: sb.crowe@gmail.com

T. Kairn · R. T. Knight · B. Hill
Genesis Cancer Care Queensland, Brisbane, QLD, Australia

J. Kenny
Epworth Radiation Oncology, Melbourne, VIC, Australia

Introduction

Intensity-modulated radiotherapy techniques (IMRT, including both IMAT and VMAT) can provide improved dose distributions in treatments of cancerous lesions that have concavities or that are close to multiple critical structures [1]. There are, however, widely recognised disadvantages to treating with unduly complex IMRT fields [2, 3]. Treatments using complex arrangements of small field segments require longer beam-on time (more MU/Gy) to deliver, resulting in increased risks of patient movement [2] and increased leakage doses to the patient [3]. Such treatments also result in increased mechanical stress, increased opportunity for treatment delivery errors and therefore an increased quality assurance workload [2, 4, 5].

Generally, the inverse-planning of IMRT treatments requires a compromise between the need to increase the complexity of the treatment plan, in order to maximise the conformation of the dose to the target; and the need to limit the complexity of the treatment plan, in order to maximise the likelihood of accurate delivery of the planned dose within a reasonable time frame [5, 6]. Contemporary radiotherapy treatment planning systems use cost functions, incorporating various treatment plan complexity metrics, for optimising the conformity and complexity of IMRT plans [7–9]. The use of similar metrics to evaluate the complexity of optimised IMRT treatment plans has been investigated by other authors [9, 10]. Such post-processing of IMRT treatment plans has the potential to provide planners with information on the complexity of a given plan, in relation to the achievable complexity of treatments to the same site [3, 9, 11].

Plan evaluation using a metric that is strongly correlated with treatment plan dosimetric accuracy could also reduce the time required for pre-treatment quality

Published online: 09 May 2014

assurance (QA) measurements and the delay before re-planning of undeliverable treatments, by identifying IMRT plans that are likely to fail their QA tests. The development of such a metric requires a broadening of the narrow definition of 'complexity' as fluence variability, used in the development of early complexity metrics [3, 11, 12]. Dosimetric studies have suggested that the potential for a given plan to be accurately delivered can also depend on the accuracy with which MLC effects are modelled by the treatment planning system, the proportion of the treatment that is delivered using low-MU segments, the size and location of individual beam apertures and the differences in area between the field apertures defined by the jaws and the (smaller) apertures defined by the MLC [4, 5, 13]. McNiven et al. [5] developed the 'modulation complexity score' (MCS) as an attempt to incorporate the effects of fluence map complexity and MLC-aperture complexity within one metric. While the MCS was shown to be capable of distinguishing between the levels of complexity required for planning treatments to different sites [5] and has shown potential for distinguishing between IMRT plans optimised with different levels of fluence smoothing [10], this metric has not been found to correlate with increasing dosimetric error for individual treatment sites [13, 14].

In this study, therefore, we attempt to identify a metric that can usefully distinguish between IMRT beams that pass and fail an array-based QA test. To this end, we re-implement the modulation index (MI) [6, 11] and the fluence map complexity (FMC) [9] fluence-complexity based metrics, which quantify the level of modulation of each beam. The sensitivities of the fluence-complexity and aperture-complexity components of the MCS [5] are investigated. Due to the difficulties associated with obtaining accurate dose calculations for small field apertures [15, 16], the potential for MLC leaf transmission to vary with field size [4] and the increased uncertainty associated with calculations of dose away from the beam's central axis [17, 18], we also evaluate novel metrics designed to be sensitive to these effects. Specifically, we have produced metrics that describe the sizes of the MLC apertures (mean field area, MFA, and small-aperture score, SAS), the proportion of closed leaves within the jaw-defined fields (closed leaf score, CLS) and the displacement of the MLC apertures, across the central axis (cross-axis score, CAS, and mean aperture displacement, MAD).

This study aims to identify a metric that can be reliably used to identify treatment plans that are unlikely to pass pre-treatment QA tests, with the goal of potentially improving the efficiency and reliability of the IMRT treatment planning and delivery process.

The initialisms used in this paper are summarised, for reference, in the Appendix.

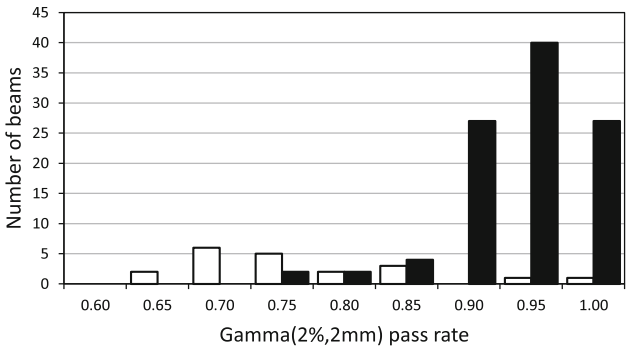

Fig. 1 Histogram of gamma pass rates for individual beams from plans that were defined as 'passed' (*black columns*) and from plans that were defined as 'failed' (*white columns*)

Methods

Treatment plans

The QA results for 122 IMRT beams from 18 prostate treatments planned and delivered over a 6 month period, using the Brainlab iPlan treatment planning system and Brainlab m3 micro-MLC (Brainlab, Feldkirchen, Germany) are included in this study. All treatments were for localised adenocarcinoma, with no nodal involvement and no metastasis. The IMRT fields had an average of 46 control points (23 beam segments) per beam, which were delivered in step-and-shoot mode. For each field, the linear accelerator's orthogonal jaws were fitted to the maximum lateral micro-MLC positions and remained stationary throughout the beam delivery.

Pre-treatment quality assurance measurements were obtained using a MapCheck2 diode array (Sun Nuclear Corporation, Melbourne, USA), positioned with its active layer at the isocentre, under 5 cm water-equivalent thickness of plastic. All beams were delivered from a zero gantry angle, but collimator rotations were preserved as per the treatment plans. Plans were passed and approved for treatment if 90 % of measurement points (above a threshold of 10 % of the maximum dose) resulted in $\gamma(2\,\%, 2\,\mathrm{mm}) < 1.0$ or if 95 % of measurement points resulted in $\gamma(3\,\%, 3\,\mathrm{mm}) < 1.0$, when compared with dose to the measurement plane calculated by the treatment planning system. Plans that did not meet these criteria were failed. (Ion chamber measurements were also obtained, to verify planned dose outputs, but these are not discussed in this study).

Figure 1 shows a histogram of the MapCheck2 γ pass rates obtained for the 122 beams used in this study. While most beams passed the 90 % threshold for the 2 %, 2 mm criteria, 8 beams passed only when the 3 %,3mm criteria was used, and 20 beams belonged to plans that failed the QA test. The range of $\gamma(2\,\%, 2\,\mathrm{mm}) < 1.0$ pass rates across these treatment plans (65–100 %) provides a useful sample for comparison with beam complexity metrics.

Treatment plans were exported from the treatment planning system, in DICOM-RT format, and analysed using the in-house TADA (Treatment and Dose Assessor) software, adapted from the existing MCDTK (Monte Carlo DICOM tool-kit) code [19, 20]. The TADA software has previously been used to retrospectively examine dosimetric quality for large numbers of patient treatments [21]. The modified TADA code allowed the batch calculation of several established and novel beam complexity metrics on a per-plan and per-beam basis, for evaluation against QA pass rates.

Mean complexity metrics for the group of plans that passed QA and the group of plans that failed QA were calculated. F-statistic values were calculated under the null hypothesis that there was no significant linear correlation between metrics and γ pass rates, as the ratio of the regression sum of squares and residual sum of squares for linear least squares regression. Corresponding p values, representing the probability of a larger F-statistic value occurring by chance, were also calculated. The Šidák correction was used to determine a significance level for each comparison from a familywise error rate, $\overline{\alpha}$, of 0.05, in order to compensate for multiple comparisons.

Established metrics

Two complexity metrics calculated using intensity fluence maps were implemented: the modulation index (MI) proposed by Webb [6] and the fluence map complexity (FMC) proposed by Llacer et al. [9]. These implementations required the generation of intensity fluence maps using the beam arrangement data contained in the RTPLAN DICOM exports. The leaf and jaw positions of each segment in the beam were used to calculate transmission on a 160×160 beam element (bixel) map representing a possible $40 \times 40\,\text{cm}^2$ field at isocentre. Where the collimators moved between beam control points the collimator positions were interpolated for 100 steps. The default leaf transmission was assumed to be 1 %, approximating values reported in the literature [22, 23]. The intensity maps were generated without consideration of collimator rotation.

The calculation of MI involved the comparison of deviations among adjacent bixels in a spectrum of the intensity fluence map to the standard deviation (σ_1) of the intensity fluence map. The calculation method presented by Nicolini et al. [11] was utilised:

$$\mathrm{MI}(F) = \int_0^F \frac{\left[Z_x(f) + Z_y(f) + Z_{xy}(f)\right]}{3}\, df \qquad (1)$$

where each Z spectra are the fraction of changes among adjacent bixels in the X, Y and X–Y directions that exceed $f\sigma_1$. An integration limit of F = 1 was used for the values presented in this work. The calculation of MI values has previously been implemented in the clinical EPID dosimetry QA software 'Epiqa' (EPIdos, Bratislava, Slovakia) [24]. It has been shown that MI has a monotonic relationship with dose conformality [6].

This FMC metric is a root sum of differences between fluence bixels (a_j) and lateral neighbouring bixels, normalised to the sum of each bixel:

$$\mathrm{FMC} = \frac{1}{\sum_j a_j} \sqrt{\sum_j \left(a_j - \lambda_k \sum_{k \in N_j} a_k \right)^2} \qquad (2)$$

where λ_k is the weight assigned to the change between neighbouring bixels (0.5 where a bixel has 2 lateral neighbours). A FMC value of 0 would indicate that there are no changes in bixel intensity between lateral bixels, while a value of 1 would imply a chequerboard pattern.

The modulation complexity score (MCS) proposed by McNiven et al. [5] was also implemented. The MCS metrics characterises modulation with two parameters: aperture area variability (AAV), the difference between leaf pair apertures for any segment compared to the maximum leaf separation in the beam; and leaf sequence variability (LSV), the variation between adjacent leaves in the same leaf bank. A high MCS value indicates little modulation of the beam intensity. The MCS metric was developed to assess plan deliverability by correlation with MapCheck gamma pass rates [5].

Novel metrics

An additional five new metrics were developed for the TADA code. These novel metrics are conceptually and computationally simpler than the established MI, FMC and MCS metrics. They were primarily designed to be sensitive to the treatment plan parameters that are most likely to compromise accurate dose calculations; small field and small segment aperture sizes [15, 16], closed MLC leaves below open linac jaws [4], and small field segments delivered from off-axis positions [17, 18].

The mean field area (MFA) was calculated as a weighted mean of the area between exposed open leaf pairs for all segments of the beam, each weighted according to the number of MU delivered

$$\mathrm{MFA}_{\mathrm{beam}} = \sum_{i=1}^{I} A_i \times \frac{\mathrm{MU}_i}{\mathrm{MU}_{\mathrm{beam}}} \qquad (3)$$

where I is the number of segments in the beam and A is the aperture area between opposing leaves.

The small aperture score (SAS) was calculated as the ratio of open leaf pairs where the aperture was less than a

defined criteria (2, 5, 10 and 20 mm in this study) to all open leaf pairs

$$\mathrm{SAS}(x)_{\mathrm{beam}} = \sum_{i=1}^{I} \frac{N(x > a > 0)_i}{N(a > 0)_i} \times \frac{\mathrm{MU}_i}{\mathrm{MU}_{\mathrm{beam}}} \quad (4)$$

where x is the aperture criteria, I is the number of segments in the beam, N is the number of leaf pairs not positioned under the jaws, and a is the aperture distance between opposing leaves.

The closed leaf score (CLS) was calculated as the ratio of closed leaf pairs to all leaf pairs inside the jaw-defined field

$$\mathrm{CLS}_{\mathrm{beam}} = \sum_{i=1}^{I} \frac{N(a \neq 0)_i}{N_i} \times \frac{\mathrm{MU}_i}{\mathrm{MU}_{\mathrm{beam}}} \quad (5)$$

where I is the number of segments in the beam, N is the number of leaf pairs not positioned under the jaws, and a is the aperture distance between opposing leaves.

The cross-axis score (CAS) was calculated as the ratio of leaf pairs where a leaf crosses the "0" position aligned with the central beam axis to all open leaf pairs

$$\mathrm{CAS}_{\mathrm{beam}} = \sum_{i=1}^{I} \frac{N(a > m)_i}{N(a > 0)_i} \times \frac{\mathrm{MU}_i}{\mathrm{MU}_{\mathrm{beam}}} \quad (6)$$

where I is the number of segments in the beam, N is the number of leaf pairs not positioned under the jaws, a is the aperture distance between opposing leaves and m is the centre of the aperture distance between opposing leaves (where 0 indicates alignment with the central beam axis).

These SAS, CLS and CAS values were calculated for each segment of the beam, and a weighted average calculated according to the number of MU delivered in each segment. A value of 1 indicates all leaf pairs inside the jaw-defined field have either an aperture less than the given criteria, are closed, or cross the axis, respectively.

The mean asymmetry distance (MAD) is the weighted average of the distance between of the centre of every open leaf pair aperture m and the central beam axis

$$\mathrm{MAD}_{\mathrm{beam}} = \sum_{i=1}^{I} \left(\sum_{j=1}^{J} |m_j| \right) \times \frac{\mathrm{MU}_i}{\mathrm{MU}_{\mathrm{beam}}} \quad (7)$$

where I is the number of segments in the beam, J is the number of leaves in each opposing bank, and m is the centre of the aperture distance between opposing leaves (where 0 indicates alignment with the central beam axis).

Results

Table 1 shows the calculated mean beam complexity metrics for the beams in plans that passed and failed QA testing. There were 4 metrics which statistics suggested could be useful in predicting MapCheck pass rates: MI, MFA, and SAS for 5 and 10 mm.

Table 1 Summary of mean plan complexity metrics for plans that passed ($\mathrm{Mean}_{\mathrm{pass}}$) and failed ($\mathrm{Mean}_{\mathrm{fail}}$) QA testing

Metric	$\mathrm{Mean}_{\mathrm{pass}}$	$\mathrm{Mean}_{\mathrm{fail}}$	F value	p value
MCS	0.37 ± 0.05	0.35 ± 0.05	0.140	0.646
AAV	0.49 ± 0.06	0.47 ± 0.06	0.097	0.756
LSV	0.75 ± 0.02	0.75 ± 0.02	0.506	0.478
MI	0.012 ± 0.002	0.015 ± 0.002	**11.397**	**0.001**
FMC	0.008 ± 0.001	0.008 ± 0.001	0.002	0.965
MFA	1300 ± 200	1200 ± 200	5.439	0.021
SAS (2 mm)	0.17 ± 0.06	0.21 ± 0.03	3.117	0.080
SAS (5 mm)	0.19 ± 0.06	0.25 ± 0.03	**9.918**	**0.002**
SAS (10 mm)	0.26 ± 0.07	0.32 ± 0.05	6.163	0.014
SAS (20 mm)	0.50 ± 0.09	0.56 ± 0.07	2.984	0.087
CLS	0.10 ± 0.04	0.12 ± 0.05	0.346	0.558
CAS	0.65 ± 0.09	0.68 ± 0.08	1.818	0.180
MAD	20 ± 4	21 ± 4	0.859	0.356

F statistic tests for significant linear relationship between the metric and γ pass rates. Critical F statistic values were 3.92 for $\alpha = 0.05$, 6.85 for $\alpha = 0.01$, and 11.38 for $\alpha = 0.001$. The p value represents the probability of a larger F statistic occurring by chance. Significant F statistic values ($\alpha = 0.0025$, after Šidák correction to experiment-wide $\bar{\alpha} = 0.05$) are shown in boldface type

Figure 2a–d respectively show the calculated MI, FMC, MCS and AAV values (that is, the established metric values) for the prostate fields examined in this study, plotted against the proportion of points that resulted in $\gamma(2\,\%, 2\,\mathrm{mm}) < 1.0$ when MapCheck2 measurement results were compared with dose planes from the treatment planning system.

The slope of the trend line for the calculated MI values is significantly different to zero ($p = 0.001$), implying correlation, while no significant correlation was observed between $\gamma(2\,\%, 2\,\mathrm{mm}) < 1.0$ and FMC, MCS and AAV values.

Figure 3a–d respectively show the calculated MFA, MAD, CAS and CLS values (that is, the novel metric values) for the prostate fields examined in this study, plotted against the proportion of points that resulted in $\gamma(2\,\%, 2\,\mathrm{mm}) < 1.0$ when MapCheck2 measurement results were compared with dose planes from the treatment planning system. Figure 3e–h show results of specifically investigating the proportion of small segment apertures that are used in each treatment (SAS scores, definitions, thresholds and false-negative rates).

The slope of the trend line for the SAS5 metric was significantly different to zero ($p \approx 0.002$). There were 2 novel metrics for which correlation with γ pass rates was

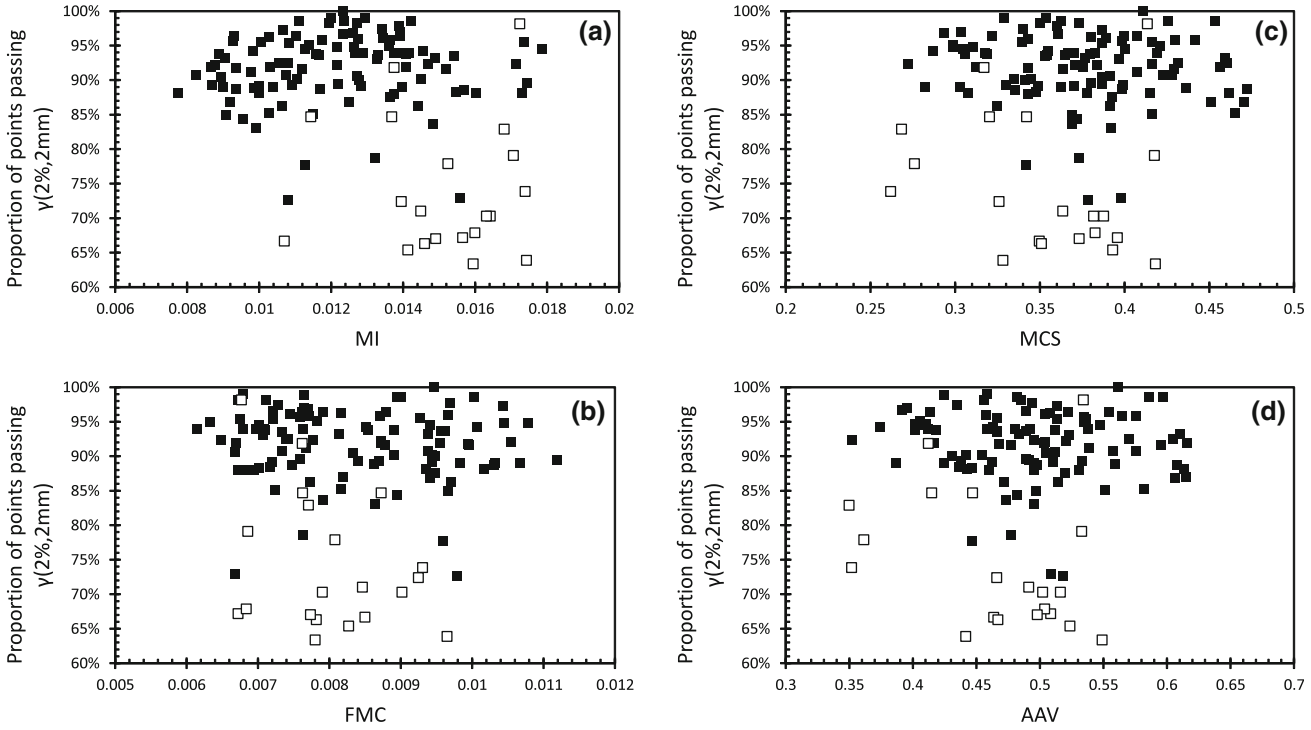

Fig. 2 a–d Treatment plan metrics evaluated using established methods (**a** MI, **b** FMC, **c** MCS and **d** AAV) plotted against QA results (proportion of points satisfying $\gamma(2\,\%, 2\,\text{mm}) < 1.0$) for beams from plans that passed their QA assessment (*filled data points*) and plans that failed their QA assessment (*open data points*). Linear regression used for trend line

potentially significant: MFA and SAS10 ($p \approx 0.02$ and $p \approx 0.01$ respectively). No significant correlation between the calculated values of the MAD, CAS, CLS, SAS2 and SAS20 metrics and γ pass rates was suggested by the calculated p values (see Table 1).

Discussion

Established metrics

Figure 2a suggests that MI is significantly correlated with plan dosimetric accuracy ($p < 0.01$ for trend line), with a 10 % decrease in MI leading to an average 20 % increase in the proportion of points satisfying $\gamma(2\,\%, 2\,\text{mm}) < 1.0$. This is supported by the significance of variation in Table 1. However, Fig. 2a also shows a strong overlap between the MI values of beams from plans that passed and failed their QA tests; there is no value of MI above which all plans have poor gamma evaluation results.

No relationship between FMC and QA result appears in the data shown in Fig. 2b, where FMC values have a small range (0.006–0.011) and show no systematic variation with the gamma evaluation results. Comparison of this result with the MI result in Fig. 2a suggests that the local variation in the intensity of beamlets is relatively consistent across these prostate plans, and that the MI result is therefore dominated by the differences in whole-beam standard deviations between the different treatment plans.

The MCS data shown in Fig. 2c is not significantly correlated to QA results. Table 1 did not suggest a significant difference between the plans that failed QA (open data points) and the plans that passed QA (filled data points), and a strong overlap between the data sets can be observed. The MCS values shown in Fig. 2c suggest that the plans evaluated here are, on average, more complex than the prostate plans previously reported by McGarry et al. [10], which were themselves more complex than the prostate plans reported by McNiven et al. [5].

Figure 2d shows AAV data that are quantitatively different but qualitatively similar to the MCS data in Fig. 2b. As noted in "Established metrics" section, the MCS for a given beam is the product of LSV and AAV. For the prostate treatments studied here, the LSV is relatively constant across all beams, with a mean value of 0.75 ± 0.02. The variation in the MCS results between beams, shown in Fig. 2b, is therefore largely determined by the variation in AAV. Figure 2c, d show results for individual beams, but MCS, LSV and AAV were also evaluated for entire treatments, producing results (not shown) similar to these individual beam results. The results of our specific examination of the effect of beam aperture size on dosimetric accuracy are discussed in the following section.

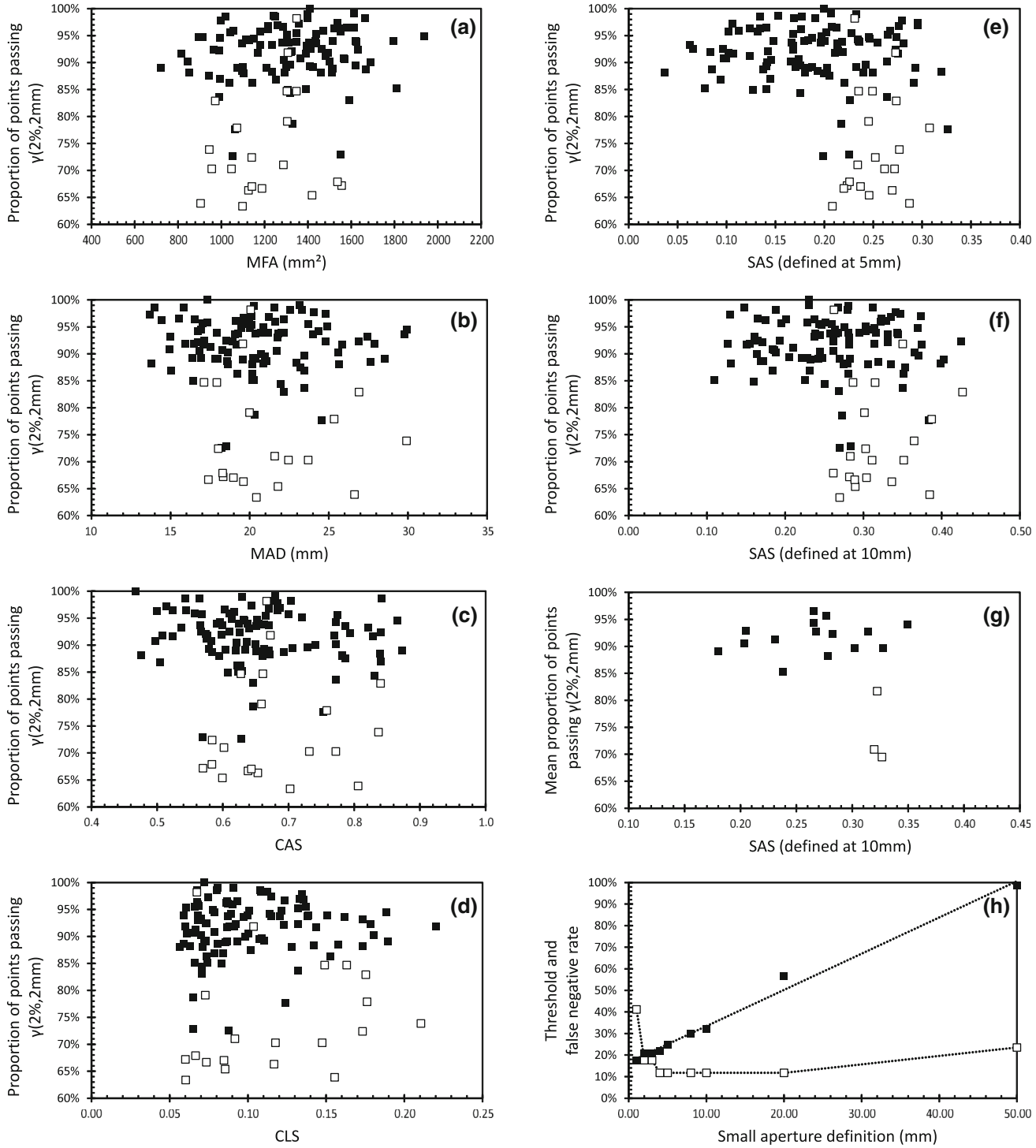

Fig. 3 **a**–**g** Treatment plan metrics evaluated using simple, novel methods (**a** MFA, **b** MAD, **c** CAS, **d** CLS, **e** SAS for each beam, where the small aperture is defined as 2 mm, **f** SAS for each beam, where the small aperture is defined as 10 mm, **g** SAS for entire treatment plans, where the small aperture is defined as 10 mm) plotted against QA results (proportion of points satisfying $\gamma(2\,\%, 2\,\text{mm}) < 1.0$) for beams from plans that passed their QA assessment (*filled data points*) and plans that failed their QA assessment (*open data points*). **h** SAS threshold value (*filled data points*) and false negative rate (*open data points*) plotted against defined small aperture size. Linear regression used for trend line

Novel metrics

Figure 3a suggests that MFA is potentially correlated with plan dosimetric accuracy ($p \approx 0.02$), with a 10 % increase in MFA leading to an average 50 % increase in the proportion of points satisfying $\gamma(2\,\%, 2\,\text{mm}) < 1.0$. This result supports the expectation that the dose from small micro-collimated IMRT segments is not as accurately predicted by the treatment planning system as the dose delivered via larger apertures [15]. However, this correlation between beam segment area and gamma evaluation result does not provide a threshold MFA value that can be used to identify plans that are likely to fail QA because, as Table 1 and Fig. 3a show, there is a strong overlap between the MFA values of beams from plans that passed and failed their QA tests.

The calculated MAD and CAS values, shown in Fig. 3b, c, are not significantly correlated to the QA pass rates, and it is not possible to unambiguously identify plans that failed their QA tests. Both metrics result in a threshold (16 mm for MAD and 0.52 for CAS) below which all plans pass their QA tests, but both metrics result in a substantial false-negative rate (more than 50 %) above this threshold. Figure 3d shows that a similar metric, quantifying the proportion of MLC leaves in each beam that are closed beneath open jaws (CLS), does not produces a useful correlation. Here, there is total overlap between the CLS ranges of the plans that passed and failed their QA tests. Evidently, all of these prostate treatments use beam segments with between 5 and 20 % of the MLC leaves closed, and this has no systematic effect on the QA outcomes.

By contrast, Fig. 3e, f show that less equivocal results can be achieved when the proportion of segment apertures defined as 'small' (SAS) is computed ($p \approx 0.002$ and $p \approx 0.01$ respectively). Figure 3e shows the result when a small aperture is defined as one where the MLC leaf separation is less than 5 mm and Fig. 3f shows the result when a small aperture is defined as one where the MLC leaf separation is less than 10 mm. All small aperture definitions used in this study resulted in similar results; increasing dosimetric accuracy is correlated with decreasing proportion of small apertures. The 10 mm result shown in Fig. 3f, is the most unequivocal of the beam analysis results and shows a threshold SAS of 0.27, below which all plans pass their QA tests. Above this threshold, there is a substantial overlap between the results for beams from plans that passed and failed their QA tests.

When the SAS is evaluated over entire plans, rather than per beam, results similar to Fig. 3g are obtained. Figure 3g shows the result when the small aperture is defined as one where the MLC leaf separation is less than 10 mm. In this figure, all plans with SAS below the threshold of 0.27 pass their QA tests. Above this threshold, only 2 of the 15 plans that passed their QA tests are grouped with the plans that failed. If the SAS, using the 10 mm small aperture definition and a threshold value of 0.27, was used to evaluate these plans before QA, it would identify all of the failed plans and 2 out of the 15 passed plans as likely to fail QA.

Figure 3h shows (as filled data points) the thresholds, SAS values below which all plans passed their QA tests, resulting from evaluating the SAS using a range of different definitions of a 'small aperture'. This threshold increases approximately linearly with increasing small aperture definition. Figure 3h also shows (as open data points) the false negative rates (number of passed plans grouped with the failed plans, above the threshold / total number of passed plans) for the different small field definitions and thresholds. The false negative rate is relatively high when the small aperture is defined as 1 mm, but decreases to a minimum when the small aperture is defined as 5–10 mm, and then increases again as the definition of the small aperture is further increased. Use of 10 mm as the 'small aperture' definition is advantageous because it falls into a broad minima in the false-negative curve; 10 mm is not so small that such apertures are contained in many segments of many beams and is not so large that it becomes less useful for identifying apertures where the treatment planning system's dose calculation is questionable.

Conclusions

This study evaluated some new and established IMRT treatment plan complexity metrics, with a view to identifying which of them might be useful for identifying likely QA failures and thereby improving the efficiency of the IMRT treatment planning and delivery process. A statistically significant correlation was identified between MI and γ pass rates. However, there was substantial overlap between the MI values of the beams that passed and failed their QA tests, limiting the utility of this metric as a treatment plan evaluation tool.

Several of the novel metrics evaluated in this study (specifically, the MAD, CAS and CLS) failed to show a statistically significant correlation with γ pass rate or provide a means to identify the likelihood of QA success, despite quantifying beam properties that were expected to affect dose calculation accuracy.

By contrast, the SAS metrics, which describe the proportion of small MLC apertures in each beam, were found to be well correlated with γ pass rates (particularly when the 'small' aperture was defined as an MLC separation of 5 mm) and provided threshold values below which all plans passed their QA tests, and above which only a small number of passing plans were identified as likely to fail.

While this outcome should be immediately useful to radiotherapy centres where prostate IMRT treatments are

planned and delivered using the Brainlab system and tested using the MapCheck2 diode array, these results also suggest avenues for further investigation at centres where IMRT treatments are delivered to other sites or using other systems. The TADA software is able to process treatment plans exported in the DICOM-RT format from a wide range of planning systems. The usefulness of fluence map complexity metrics may vary between treatment sites and inverse-planning optimisation systems, with the variability of each metric depending on the degree of local (FMC) or global (MI) beam fluence modulation. Further development or investigation of novel treatment plan complexity metrics should concentrate on metrics that describe the beam aperture size (MFA, SAS) rather than the beam aperture position (MAS, CAS).

Appendix: Summary of initialisms

- AAV: Aperture area variability
- CAS: Cross axis score
- CLS: Closed leaf score
- FMC: Fluence map complexity
- LSV: Leaf sequence variability
- MAD: Mean aperture displacement
- MCDTK: Monte Carlo DICOM tool-kit
- MCS: Modulation complexity score
- MFA: Mean field area
- MI: Modulation index
- SAS: Small aperture score
- TADA: Treatment and dose assessor

References

1. Purdy JA, Boyer AL, Butler EB, Dipetrillo TA, Engler MJ, Fraass B, Grant W III, Ling CC, Low DA, Mackie TR, Mohan R, Roach M, Rosenman JG, Verhey LJ, Wong JW, Cumberlin RL, Stone H, Palta JR (2001) Intensity-modulated radiotherapy: current status and issues of interest. Int J Radiat Oncol Biol Phys 51(4):880–914
2. Lee MT, Purdie TG, Eccles CL, Sharpe MB, Dawson LA (2010) Comparison of simple and complex liver intensity modulated radiotherapy. Radiat Oncol 5:115
3. Nauta M, Villarreal-Barajas JE, Tambasco M (2011) Fractal analysis for assessing the level of modulation of IMRT fields. Med Phys 38(10):5385–5393
4. Fenoglietto P, Laliberé B, Aillères N, Riou O, Dubois JB, Azria D (2011) Eight years of IMRT quality assurance with ionization chambers and film dosimetry: experience of the Montpellier Comprehensive Cancer Center. Radiat Oncol 6:85
5. McNiven AL, Sharpe MB, Purdie TG (2010) A new metric for assessing IMRT modulation complexity and plan deliverability. Med Phys 37(2):505–515
6. Webb S (2003) Use of a quantitative index of beam modulation to characterize dose conformality: illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy. Phys Med Biol 48(14):2051–2062
7. Spirou SV, Chui CS (1998) A gradient inverse planning algorithm with dose-volume constraints. Med Phys 25(3):321–333
8. Bortfeld T, Bürkelbach J, Boesecke R, Schlegel W (1990) Methods of image reconstruction from projections applied to conformation radiotherapy. Phys Med Biol 35(10):1423–1434
9. Llacer J, Solberg TD, Promberger C (2001) Comparative behaviour of the dynamically penalized likelihood algorithm in inverse radiation therapy planning. Phys Med Biol 46(10): 2637–2663
10. McGarry CK, Chinneck CD, O'Toole MM, O'Sullivan JM, Prise KM, Hounsell AR (2011) Assessing software upgrades, plan properties and patient geometry using intensity modulated radiation therapy (IMRT) complexity metrics. Med Phys 38(4):2027–2034
11. Nicolini G, Fogliata A, Vanetti E, Clivio A, Ammazzalorso F, Cozzi L (2007) What is an acceptably smoothed fluence? Dosimetric and delivery considerations for dynamic sliding window IMRT. Radiat Oncol 2:42
12. Kim T, Zhu L, Suh T-S, Geneser S, Meng B, Xing L (2011) Inverse planning for IMRT with nonuniform beam profiles using total-variation regularization (TVR). Med Phys 38(1):57–66
13. Tonigan J, Kry S, Dong L, Purdie T, White R, Ibbott G, Followill D (2011) Does IMRT treatment plan complexity or mismatched dosimetry data contribute to dose delivery errors detected using an IMRT H&N quality assurance phantom? Med Phys 38(6): 3804
14. Tonigan JR (2011) Evaluation of intensity modulated radiation therapy (IMRT) delivery error due to IMRT treatment plan complexity and improperly matched dosimetry data. Master's Thesis, University of Texas
15. Kairn T, Hardcastle N, Kenny J, Meldrum R, Tomé WA, Aland T (2011) EBT2 radiochromic film for quality assurance of complex IMRT treatments of the prostate: micro-collimated IMRT, RapidArc, and TomoTherapy. Australas Phys Eng Sci Med 34(3):333–343
16. Kairn T, Crowe S, Kenny J, Trapp JV (2011) Investigation of stereotactic radiotherapy dose using dosimetry film and Monte Carlo simulations. Radiat Meas 46(12):1985–1988
17. Ahnesjö A, Aspradakis MM (1999) Dose calculations for external photon beams in radiotherapy. Phys Med Biol 44(11):R99–R155
18. Brainlab AG (2010) Brainlab physics technical reference guide. Revision 1:2
19. Crowe S, Kairn T, Fielding AL (2009) The development of a Monte Carlo system to verify radiotherapy treatment dose calculations. Radiother Oncol 92(Supp 1):S71
20. Crowe SB, Kairn T, Trapp JV, Fielding AL (2013) Experimental evaluation of MCDTK, the Monte Carlo DICOM ToolKit. IFMBE Proc 39:1807–1810
21. Crowe SB, Kairn T, Middlebrook N, Bill B, Christie DRH, Knight RT, Kenny J, Langton CM, Trapp JV (2013) Retrospective evaluation of dosimetric quality for prostate carcinomas treated with 3D conformal, intensity-modulated and volumetric-modulated arc radiotherapy. J Med Radiat Sci 60(4):131–138
22. Arnfield MR, Siebers JV, Kim JO, Wu Q, Keall PJ, Mohan R (2000) A method for determining multileaf collimator transmission and scatter for dynamic intensity modulated radiotherapy. Med Phys 27(10):2231–2241
23. Cosgrove VP, Jahn U, Pfaender M, Bauer S, Budach V, Wurm RE (1999) Commissioning of a micro multi-leaf collimator and planning system for stereotactic radiosurgery. Radiother Oncol 50(3):325–336
24. Nicolini G, Vanetti E, Clivio A, Fogliata A, Korreman S, Bocanek J, Cozzi L (2008) The GLAaS algorithm for portal dosimetry and quality assurance of RapidArc, an intensity modulated rotational therapy. Radiat Oncol 3:24