IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology

Phys. Med. Biol. **60** (2015) 2587–2601 doi:10.1088/0031-9155/60/6/2587

Examination of the properties of IMRT and VMAT beams and evaluation against pre-treatment quality assurance results

S B Crowe[1,2], T Kairn[2,3], N Middlebrook[3], B Sutherland[3], B Hill[3], J Kenny[4], C M Langton[2] and J V Trapp[2]

[1] Royal Brisbane and Women's Hospital, Butterfield Street, Herston QLD 4006, Herston, Australia
[2] Queensland University of Technology, 2 George Street, Brisbane QLD 4000, Australia
[3] Genesis CancerCare Queensland, 1/40 Chasely Street, Auchenflower QLD 4000, Australia
[4] Epworth Radiation Oncology, 89 Bridge Road, Richmond VIC 3121, Australia

E-mail: sb.crowe@gmail.com

Received 23 October 2014, revised 8 February 2015
Accepted for publication 12 February 2015
Published 11 March 2015

Abstract

This study aimed to provide a detailed evaluation and comparison of a range of modulated beam evaluation metrics, in terms of their correlation with QA testing results and their variation between treatment sites, for a large number of treatments. Ten metrics including the modulation index (MI), fluence map complexity, modulation complexity score (MCS), mean aperture displacement (MAD) and small aperture score (SAS) were evaluated for 546 beams from 122 intensity modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT) treatment plans targeting the anus, rectum, endometrium, brain, head and neck and prostate. The calculated sets of metrics were evaluated in terms of their relationships to each other and their correlation with the results of electronic portal imaging based quality assurance (QA) evaluations of the treatment beams. Evaluation of the MI, MAD and SAS suggested that beams used in treatments of the anus, rectum, head and neck were more complex than the prostate and brain treatment beams. Seven of the ten beam complexity metrics were found to be strongly correlated with the results from QA testing of the IMRT beams ($p < 0.00008$). For example, values of SAS (with multileaf collimator apertures narrower than 10 mm defined as 'small') less than 0.2 also identified QA passing IMRT beams with 100% specificity. However, few of the metrics are correlated with the results from QA testing of the VMAT beams, whether they were evaluated as whole 360° arcs or as 60° sub-arcs. Select evaluation of beam complexity metrics

(at least MI, MCS and SAS) is therefore recommended, as an intermediate step in the IMRT QA chain. Such evaluation may also be useful as a means of periodically reviewing VMAT planning or optimiser performance.

Keywords: radiotherapy, intensity-modulated radiotherapy, volumetric-modulated arc therapy, electronic portal imaging, quality assurance, beam complexity

(Some figures may appear in colour only in the online journal)

1. Introduction

Inverse-planned modulated radiotherapy treatments (including intensity modulated radiotherapy, IMRT, and volumetric modulated arc therapy, VMAT) are difficult to evaluate and interpret, arising from optimisation algorithms that attempt to balance conflicting dosimetric goals despite unavoidable geometric constraints (Webb 2003). Based on user-accessible variables in the treatment planning system, it can be difficult to predict the likely success or failure of a given treatment plan when submitted to pre-treatment quality assurance (QA) testing. It can be similarly difficult to identify the reasons for QA failures.

The number of MU per beam or the total MU per Gy delivered can be used as a surrogate for beam complexity or a simplistic indicator of the presence of excessive numbers of small apertures or closed multileaf collimator (MLC) leaves in the treatment plan (Tonigan 2011, Tonigan *et al* 2011, Miura *et al* 2014), but there remains a manifest need for users to be able to directly and independently evaluate plan deliverability prior to QA testing.

This need has been answered, in recent years, by the development of a small number of metrics for assessing modulated plan complexity, deliverability and dose calculation accuracy; some of these metrics have been correlated with QA results via small studies (McNiven *et al* 2010, Jørgensen *et al* 2011, McGarry *et al* 2011, Tonigan 2011, Tonigan *et al* 2011, Crowe *et al* 2014, Kairn *et al* 2014).

For example, McNiven *et al* (2010) developed the modulation complexity score (MCS), which analyses MLC positions to provide an assessment of plan complexity and deliverability. This metric has been shown to be poorly correlated with phantom dosimetry QA results when head and neck plans are evaluated (Tonigan 2011, Tonigan *et al* 2011), while also clearly distinguishing between the levels of complexity involved in treating different anatomical sites (prostate compared to head and neck) (McGarry *et al* 2011).

Similarly, Kairn *et al* (2014) developed a number of metrics to quantify specific parameters expected to lead to dose calculation inaccuracy in the Brainlab iPlan treatment planning system (Brainlab, Feldkirchen, Germany). While some of these metrics produced ambiguous or inconclusive results, the small aperture score (a calculation of the proportion of the beam that is delivered via MLC apertures smaller than a given threshold) was shown to correlate with QA data obtained using a low-resolution diode array system for a small sample of prostate and cranial IMRT treatments (Kairn *et al* 2014).

Recently, the modulation index (MI), developed by Webb (2003) as a means to introduce a weighting against fluence complexity into the IMRT inverse planning cost function, has also been shown to be correlated with the gamma agreement index (GAI) achieved by IMRT beams planned for delivery in prostate treatments, using the Brainlab m3 microMLC (Crowe *et al* 2014).

To date, however, a comprehensive evaluation of a broad range of metrics, applied to a large number of beams designed for treating different treatment sites has been lacking from

Table 1. Distribution of treatment plan modalities and anatomical sites.

Treatment Site	# IMRT Plans	# IMRT Beams	# VMAT Plans	# VMAT Beams
Anus/Rectum	6	79	5	12
Endometrium	3	35	2	4
Brain	10	56	12	24
Head &Neck	3	22	11	24
Prostate	30	210	40	80
Total	52	402	70	144

the literature. Such an evaluation has the potential to identify those metrics that are able to provide meaningful information about modulated treatment beams, to evaluate possible inter-relationships between the metrics themselves, and to establish which metrics behave consistently regardless of anatomical site treated and which metrics vary between treatment sites. This study therefore aims to provide a detailed examination of beam complexity, deliverability and accuracy metrics, using 546 beams from 122 IMRT and VMAT treatment plans targeting the anus, rectum, endometrium, brain, head and neck and prostate. The information produced by this study may thereafter be used to inform treatment planning, to guide QA testing or as a QA tool in itself.

2. Methods

2.1. Treatment plans and quality assurance

This study has evaluated 402 beams from 52 IMRT treatment plans for five anatomical sites as well as 144 beams from 70 VMAT treatment plans for the same sites, as detailed in table 1. All treatments, using both modalities, were planned over the same 11 month period. The treatments were planned using the Varian Eclipse treatment planning system (version 11), using the AAA dose calculation algorithm, and delivered using a Varian iX linear accelerator, with a Millennium 120-leaf MLC (Varian Medical Systems, Palo Alto, USA). The VMAT plans used the Varian RapidArc system, with a dose rate of 600 MU min^{-1} and with MLC leaves moving back and forth across each field. The IMRT plans used the sliding-window technique, with a dose rate of 300 MU min^{-1} and with MLC leaves moving in one direction across each field.

Routine pre-treatment quality assurance checks were performed for all treatments listed in table 1 using the Epiqa system (EPIdos, Bratislava, Slovakia), using the GLAaS dosimetric method (Nicolini *et al* 2008). The Epiqa system has been detailed by Nicolini *et al* (2008) and Fogliata *et al* (2011), and the sensitivity of the system to introduced errors explored by Fredh *et al* (2013). This system involves the acquisition of portal images of each treatment beam, delivered using its planned collimator and gantry angles (including the whole 360° rotations in the VMAT plans), using the Varian aS1000 EPID (Varian Medical Systems, Palo Alto, USA), which has a 0.39×0.39 mm^2 spatial resolution.

During the data acquisition period, the stability of both the linac system (including MLC performance) and the EPID used to acquire the QA images was verified through the delivery of a set of four static MLC fields, two VMAT arcs and four IMRT beams of varying complexity. For reasons unrelated to this study, the four IMRT beams were delivered twice, at dose rates of 300 MU min^{-1} and 600 MU min^{-1}, allowing the consistency of IMRT delivery at different dose rates to be established.

In order to quantify the differences between the EPID images and the corresponding dose planes exported from the Eclipse treatment planning system, the Epiqa software was used

to calculate gamma agreement indices (GAIs), the percentage of points passing a gamma evaluation ($\gamma < 1.0$) (Low and Dempsey 2003, Wendling *et al* 2007) using standard IMRT quality assurance criteria, γ (3%, 3 mm) (Ezzell *et al* 2009, ICRUM 2010). IMRT treatments are approved for delivery in this clinic where the mean GAI over all beams is >90% and no beam has a GAI < 85% . VMAT treatments are approved for delivery where no beam has a GAI < 90% . Treatments failing to meet the acceptance criteria were replanned. The gamma evaluation was performed on the dose values recorded within the jaw collimated field.

The gamma evaluations for the plans presented here were repeated with tighter gamma evaluation criteria, γ (2%, 2 mm). This re-evaluation was intended to provide information on the effect of QA test stringency on the spread of the beam metric data. Linear regression and t-tests were used to determine the nature and significance of the correlation between GAI for γ (3%, 3 mm) and γ (2%, 2 mm). Since the results of this work showed a linear relationship between the GAIs evaluated using the two different tolerance levels (see section 4.3), only the more clinically relevant GAI, γ (3%,3 mm), has been plotted against the metric results.

Centres generally perform VMAT quality assurance checks by delivering the whole 360° rotations, but it is also possible to divide the arc into sub-arc segments and evaluate these separately (Nicolini *et al* 2008, Ravkilde *et al* 2013). Quality assurance checks have been repeated here for a representative sample of 32 arcs (with varying GAI) each divided into 60° sub-arcs. The GAIs for these beams have been evaluated against corresponding whole (i.e. 360°) arc values.

The GAI for each beam was compared against a range of calculated metrics designed to describe the complexity, deliverability and accuracy of the beam. These metrics were calculated using the in-house Treatment and Dose Assessor (TADA) software, used previously in the examination of quality assurance using diode array measurements (Crowe *et al* 2014, Kairn *et al* 2014) and the retrospective evaluation of dosimetric quality in prostate treatments (Crowe *et al* 2013). The significance of correlation between GAI and these metrics was evaluated using *F*-tests.

2.2. Complexity metrics

The TADA software uses beam arrangement data to generate intensity fluence maps representing a 40×40 cm^2 field at isocentre, with 160×160 beam elements (bixels). This fluence map is used in the calculation of two established complexity metrics: the FMC and MI.

The FMC metric, proposed by Llacer *et al* (2001) evaluates deviations between adjacent bixels relative to the sum of each bixel. The MI metric, proposed by Webb (2003), evaluates deviations between adjacent bixels relative to the standard deviation of the intensity fluence map. The TADA software uses the Nicolini *et al* (2007) implementation of MI calculation (Crowe *et al* 2014).

2.3. Deliverability metrics

The MCS was proposed by McNiven *et al* (2010) as providing a means of quantifying the deliverability of modulated treatment plans by directly evaluating the complexity of the positioning of MLC leaves, rather than the complexity of the fluence map they are positioned to deliver. The MCS combines an aperture area variability (AAV) score with a leaf sequence variability (LSV) score, to provide an indication of the planned mobility of individual MLC leaves during the delivery of the treatment.

2.4. Accuracy metrics

A set of treatment planning system dose calculation accuracy metrics were proposed by Kairn *et al* (2014), specifically for use with the pencil beam dose calculation algorithm available in the Brainlab iPlan treatment planning system (Brainlab, Feldkrichen, Germany). These metrics were designed to quantify the parameters identified as most likely to compromise accurate dose calculations; small field and small segment aperture sizes (Kairn *et al* 2011a, 2011b), closed MLC leaves below open linac jaws (Fenoglietto *et al* 2011), and small field segments delivered from off-axis positions (Ahnesjö and Aspradakis 1999, Brainlab AG 2010).

Specifically, Kairn *et al* (2014) investigated: the mean field area (MFA), an average of field area weighted according to the segment monitor units; the mean aperture displacement (MAD), the mean lateral displacement (away from the central axis) of the centre of the opening between each pair of MLC leaves; the cross-axis score (CAS), the proportion of MLC leaves within the jaw aperture that cross the central axis; the closed leaf score (CLS), the proportion of MLC leaf pairs within the jaw aperture that are entirely closed; and the small aperture score (SAS), the proportion of open MLC leaf pairs that are separated by less than a given threshold difference, in each beam. The evaluation of these metrics and their use with treatment plans from the Brainlab iPlan treatment planning system is further detailed in Crowe *et al* (2014). A list of these initialisms is provided as an appendix.

For the purposes of the current study, the MFA, MAD, CAS, CLS and SAS metrics were instead applied to treatments planned using the AAA algorithm in the Varian Eclipse treatment planning system.

3. Results

The results of comparing the GAI values from QA testing with the treatment plan complexity, deliverability and accuracy metrics are shown in figures 1–3. Figure 1 shows the fluence map metric MI, figure 2 shows the leaf positioning metric MCS, and figure 3 shows aperture based metrics CAS, MAD and SAS calculated using a 10 mm threshold.

In figures 1(*a*)–(*d*) the QA results are shown for the IMRT and VMAT beams, plotted against beam fluence complexity, represented as MI. In all four figures, MI shows an overall trend of decreasing with increasing GAI, indicating that QA higher pass rates are achieved by less complex beams, although this trend is statistically significant for the IMRT beams only (see table 2). Data shown in figures 2(*a*)–(*c*) suggest that MI can distinguish between treatments for different anatomical sites, with anus/rectum and endometrium treatment beams delivering generally more complex fluences than the prostate and brain treatment beams.

Figures 2(*c*) and (*d*) show the AAV, one of the two components of the MCS, plotted against MCS. The linear relationship between these data indicates that the MCS is largely determined by MLC aperture variability, with interdigitation of leaves (described by LSV, the other component of the MCS) varying more weakly between beams.

In contrast to the fluence map based results of figure 1, the leaf positioning MCS results shown in figures 2(*a*) and (*b*) do not clearly distinguish between treatments for different anatomical sites. There is a statistically significant trend (see table 2), with an increasing MCS (a decreasing complexity) correlated with a increasing gamma pass percentage.

Figure 3 shows that there are general trends towards decreasing CAS and decreasing MAD with increasing GAI, as shown in figures 3(*a*) and (*b*), suggesting that plans with less asymmetric MLC apertures and fewer exposed leaves crossing the central axis result in more accurate delivery of the planned doses. Figure 3(*c*) confirms this geometric relationship between the CAS and MAD.

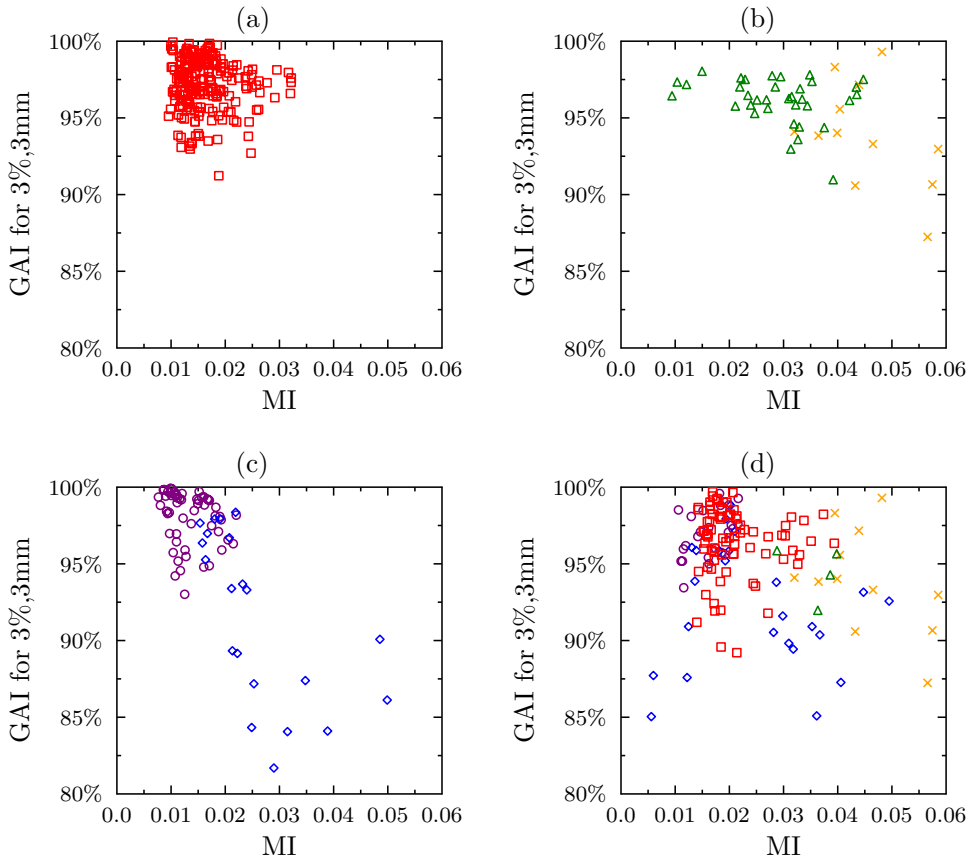

Figure 1. Relationship between GAI calculated using acceptance criteria of 3%, 3 mm and MI for (*a*)–(*c*) IMRT and (*d*) VMAT beams according to treatment site: anus (orange cross markers), brain (purple circle markers), endometrium (green triangle markers), head and neck (blue diamond markers), and prostate (red square markers).

While average MLC aperture sizes are generally larger for the VMAT beams than for the IMRT beams, especially for anus and rectum treatments, the mean field area metric has no significant correlation with QA results (see table 2). There are, however, statistically significant correlations between the proportions of each beam that are comprised of small apertures for both IMRT beams (when 'small' is defined as either narrower than 1 mm, less than 5 mm or less than 10 mm) and VMAT beams (when 'small' is defined as narrower than 1 mm). In these treatments, decreasing small aperture scores are correlated (see table 2) with increasing GAIs, as exemplified in figure 3(*d*).

The effect that the proportion of small apertures in each beam has on the modulation and complexity of IMRT beams is shown in figures 4(*a*) and (*b*). Figure 4(*a*) suggests an inconsistent relationship between SAS and MI, with beam modulation increasing with increasing proportion of small apertures for prostate, head and neck and anus treatment beams only, while an inverse relationship is apparent in the results for the endometrium treatment beams. Figure 4(*b*) illustrates a strong inverse relationship between SAS and MCS, with the beams with more complex MLC arrangements (lower MCS values) being comprised of larger proportions of small apertures (higher SAS values). Figures 4(*c*) and (*d*) indicate that this relationship also holds for the VMAT beams, even when they are evaluated as 60° sub-arcs.

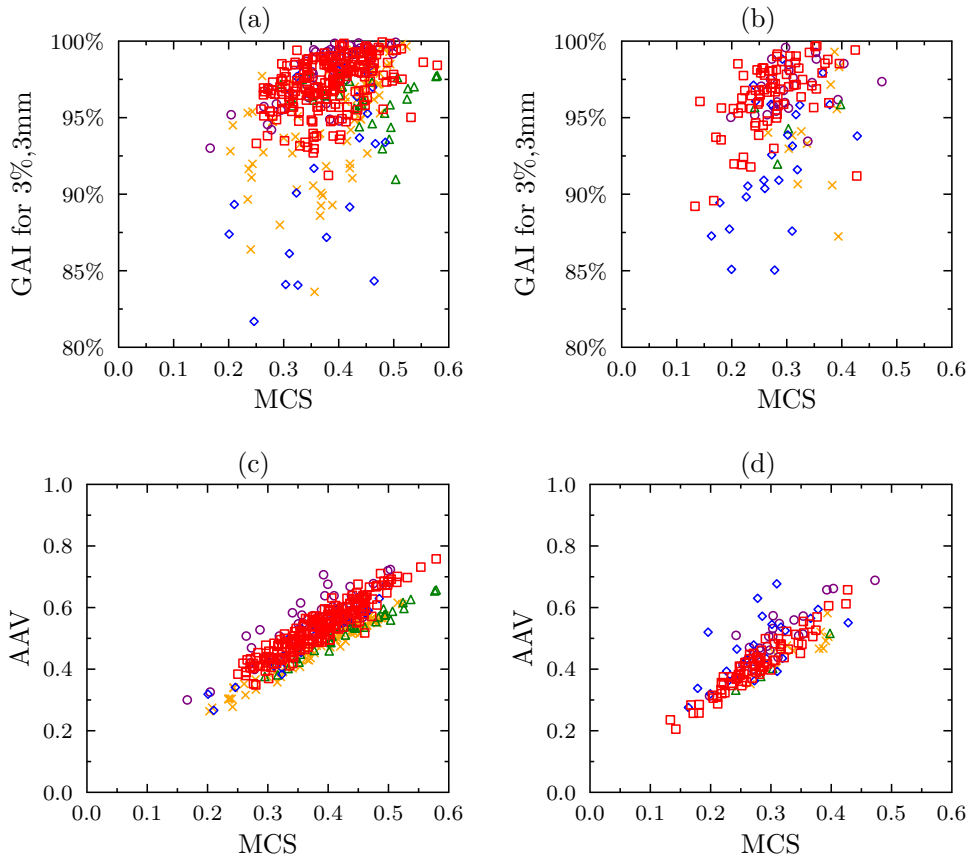

Figure 2. Relationship between GAI calculated using acceptance criteria of 3%, 3 mm and MCS for (*a*) IMRT and (*b*) VMAT beams, and relationship between AAV and MCS for (*c*) IMRT and (*d*) VMAT beams, according to treatment site: anus (orange cross markers), brain (purple circle markers), endometrium (green triangle markers), head and neck (blue diamond markers), and prostate (red square markers).

Table 2 quantifies the significance of correlations between calculated metrics and the QA data by listing *F*-test statistics. Data in table 2 indicate that while the QA results for the IMRT beams are significantly correlated with the beam complexity (MI) and deliverability (MCS, AAV) as well as the proportion of small and asymmetric MLC apertures in the beam (SAS, MAD, CAS), the QA results for the VMAT beams are not strongly correlated with any of the metrics evaluated except the MCS. The VMAT QA results' apparent insensitivity to beam modulation and composition holds even when the arcs are evaluated as series of 60° sub-arcs.

The relationship between whole 360° rotation GAIs and the GAI for 60° sub-arcs is presented in figure 5 for 3%, 3 mm and 2%, 2 mm. Over each set of data, the agreement with measurements is improved when the entire arc is delivered, by 2.5% and 5.3% when comparing against the mean GAI of the sub-arcs (at 3%, 3 mm and 2%, 2 mm). The data is summarised in table 3.

The relationship between GAI using acceptance criteria of 3%, 3 mm and 2%, 2 mm is shown in figure 6 and table 4. These results suggest that, when an evaluation of the agreement between the planned dose and the dose measured in a 2D plane within a homogeneous phantom is completed using 2%,2 mm criteria, the resulting GAI are generally (with few

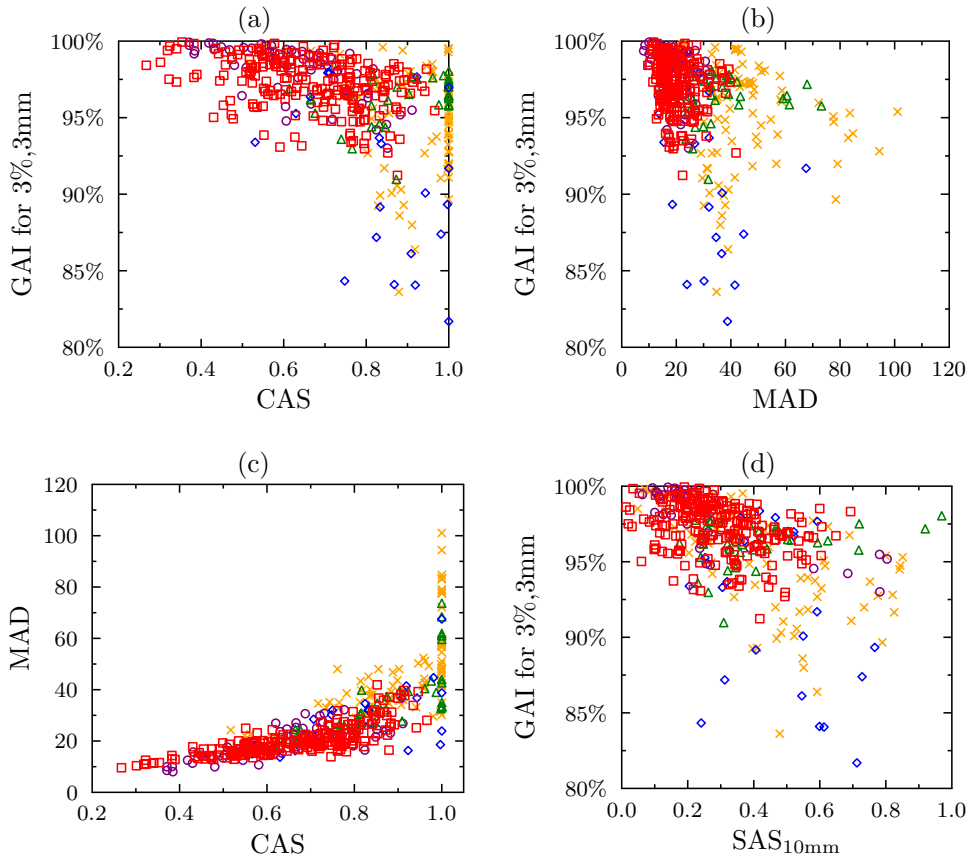

Figure 3. (*a*) Relationship between GAI calculated using acceptance criteria of 3%, 3 mm and CAS for IMRT beams. (*b*) Relationship between GAI calculated using acceptance criteria of 3%, 3 mm and MAD for IMRT beams. (*c*) Relationship between MAD and CAS for IMRT beams. (*d*) Relationship between GAI calculated using acceptance criteria of 3%, 3 mm and SAS calculated with a 10 mm threshold for IMRT beams. Data is labelled according to treatment site: anus (orange cross markers), brain (purple circle markers), endometrium (green triangle markers), head and neck (blue diamond markers), and prostate (red square markers).

exceptions) close to 8% lower than the GAI resulting from an evaluation using 3%, 3 mm criteria. In other words, for the treatment plans examined in this study, the use of a 90% pass rate for QA tests using 3%, 3 mm criteria is approximately equivalent to the use of 82% pass rate for QA tests using 2%, 2 mm criteria.

4. Discussion

4.1. Effects of anatomical site

Comparison of most of the results for different anatomical sites indicates that the anus, rectum, head and neck treatment beams were generally more complex, with modulated fluences produced by combining larger numbers of asymmetric field apertures, than the prostate endometrium and brain treatment beams. Specifically, data in figures 3(*a*) and (*b*) show that for

Table 2. Summary of *F* statistic tests for significant linear relationships between metric values and GAI for beams.

Metric	VMAT 3%, 3 mm	VMAT 2%, 2 mm	Sub-arc 3%, 3 mm	Sub-arc 2%, 2 mm	IMRT 3%, 3 mm	IMRT 2%, 2 mm
# Plans	70	70	19	19	52	52
# Beams	144	144	207	207	402	402
MU	ns	ns	ns	ns	***	***
MFA	ns	ns	ns	ns	ns	ns
AAV	ns	ns	ns	ns	***	***
CLS	ns	ns	ns	ns	ns	ns
CAS	ns	ns	ns	ns	***	***
FMC	ns	ns	***	**	ns	ns
MAD	**	***	ns	ns	***	***
MCS	**	ns	***	***	***	**
MI	*	**	ns	ns	***	***
SAS1	*	*	*	*	***	***
SAS5	ns	ns	ns	ns	***	***
SAS10	ns	ns	ns	ns	***	***

Note: Significance is evaluated against Šidák corrected α values of approximately 0.004 (*), 0.0008 (**) and 0.00008 (***), corresponding to experiment-wide α values of 0.05, 0.01 and 0.001 respectively.

the treatment plans used in this study, the beams designed to treat larger volumes use larger proportions of MLC leaves that block the central axis to produce beam apertures that are displaced off-axis by greater distances than the beams designed to treat smaller volumes. IMRT beams from the endometrium, anus and head and neck treatments were found to produce most of the high SAS results (see figure 3(*d*)), most of the high MI results (see figures 1(*b*) and (*c*)) and many of the low AAV results (see figure 2(*c*)) seen in this study.

The apparent stratification of the MCS results for different anatomical sites (especially for the IMRT beams) shown in figures 2(*c*) and (*d*) and 4(*b*)–(*d*) similarly suggest an increase in the complexity of MLC arrangements for increasing treatment volume (from the beams from the prostate plans to the beams from the brain plans to the beams from the anus and endometrium plans), which reflect the differences between MCS values from head and neck plans and prostate plans that were observed in the original description of the MCS by McNiven *et al* (2010) and the more recent phantom study by McGarry *et al* (2011).

In this study, all of the IMRT beams that failed their QA tests (GAI < 90%) came from treatment plans for the anus and rectum and head and neck regions, where large targets including lymph nodes are treated using relatively large and sometimes split radiation fields. Data in figure 3(*a*) show that all of these failing beams involved MLC leaves that were closed across the central axis more than they were open away from the central axis (with cross axis scores substantially greater than 0.5). Data in figure 1(*a*) and figures 3(*a*) and (*b*) show that the failing anus and rectum treatment beams are quantitatively different (more modulated, with more asymmetric MLC positions and more closed MLC leaves per beam) from the passing anus and rectum treatment beams.

Overall, anus, rectum, head and neck treatments contributed the majority of VMAT arcs that failed their QA tests, suggesting that the increased complexity of the treatments planned for larger target volumes, with correspondingly decreased proximity to neighbouring organs at risk, unavoidably leads the treatment planning systems optimisation engine to produce modulated beam fluences that are more difficult for the treatment delivery system to reproduce than

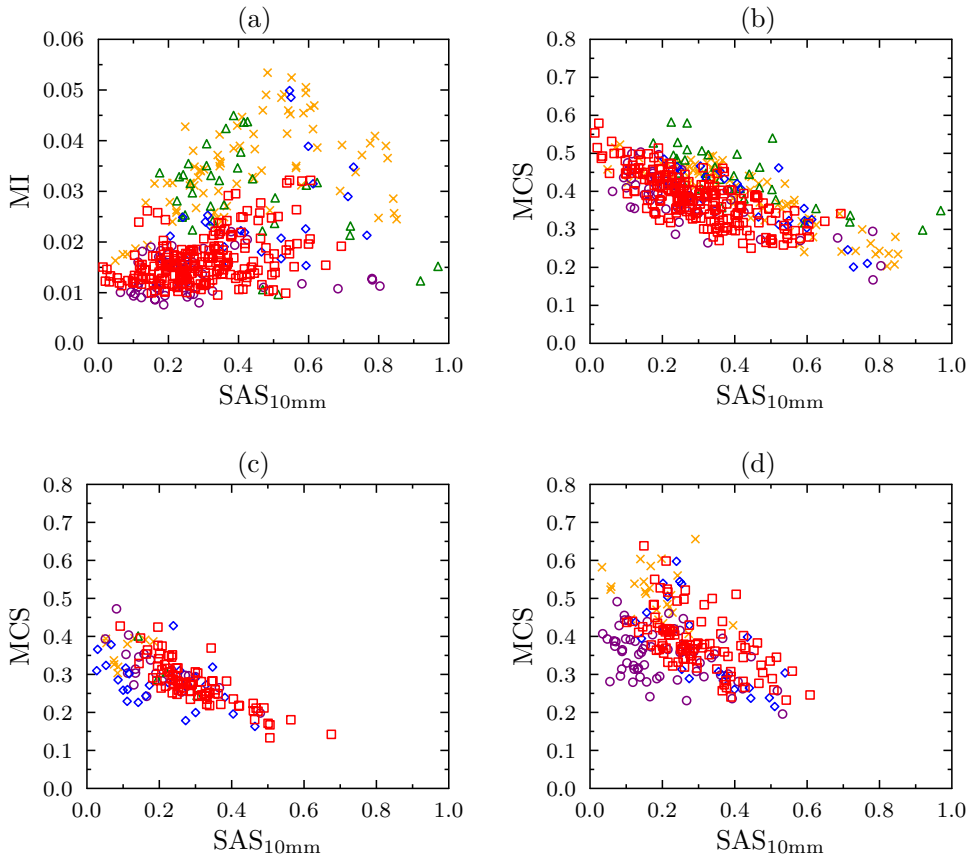

Figure 4. (*a*) Relationship between MI and SAS for IMRT beams. (*b*) Relationship between MCS and SAS for IMRT beams. (*c*) Relationship between MCS and SAS for VMAT beams evaluated as entire 360° arcs. (*d*) Relationship between MCS and SAS for VMAT beams evaluated as 60° sub-arcs. SAS has been calculated using a 10 mm threshold. Data is labelled according to treatment site: anus (orange cross markers), brain (purple circle markers), endometrium (green triangle markers), head and neck (blue diamond markers), and prostate (red square markers).

the beam fluences planned for smaller treatment volumes with fewer or less highly prioritised organs at risk.

4.2. Effects of treatment modality

Data in table 2 suggest that many of the complexity, deliverability and calculation accuracy metrics investigated in this study are strongly correlated with GAI results from QA testing of the IMRT beams, but few of the metrics are correlated with GAI results from QA testing of the VMAT beams, whether they are evaluated as whole 360° arcs or as 60° sub-arcs.

For example, for the IMRT beams, strong correlations between small aperture scores and GAI results indicates that beams with fewer small MLC apertures are more accurately delivered than beams comprised of larger proportions of small apertures. This result, obtained from the analysis of 402 beams from treatments for five different anatomical sites, verified using a high-resolution EPID-based pre-treatment QA system, provides broad confirmation of the

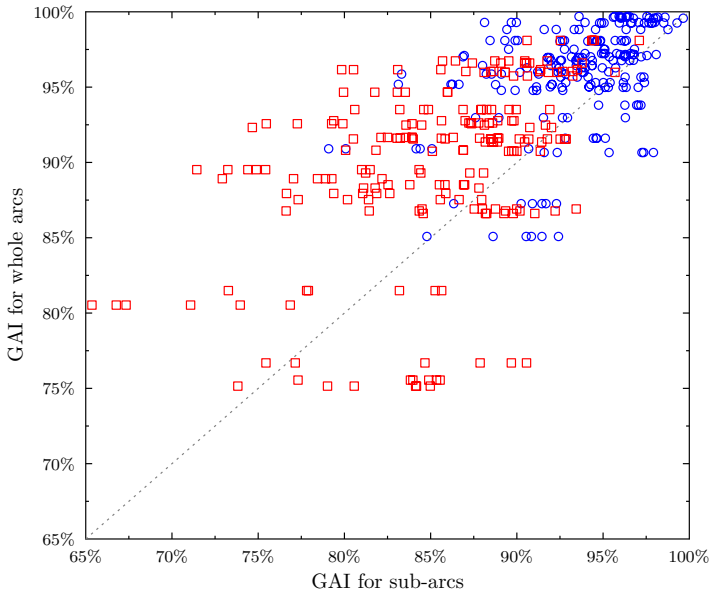

Figure 5. (*a*) Relationship between GAI calculated for whole arcs and GAI calculated for corresponding sub-arcs, for both 3%, 3 mm (blue circle markers) and 2%, 2 mm (red square markers) criteria. Dashed line indicates 1 : 1 relationship. Linear coefficient is 1.025 for 3%, 3 mm data and 1.050 for 2%, 2 mm data (assuming zero constant).

Table 3. GAI values calculated for whole arcs and corresponding sub-arc GAI sets, for both 3%, 3 mm and 2%, 2 mm.

	3%, 3 mm	2%, 2 mm
Whole arc (360°) GAI	95.9%	90.0%
Sub-arc (60°) maximum GAI	96.0%	89.2%
Significance	ns	ns
Sub-arc (60°) mean GAI	93.5%	85.4%
Significance	**	***
Sub-arc (60°) minimum GAI	90.6%	80.8%
Significance	***	***

Note: Significance of difference is evaluated between whole and divided sub-arc pairings using a two-tailed Welch's *t*-test.

results of a previous study that used only 151 beams, from prostate and cranial treatments only, verified using measurements made with a low-resolution diode array (Kairn *et al* 2014).

For the IMRT beams used in this work, the small aperture score potentially provides useful thresholds for predicting likely QA outcomes; all IMRT beams pass their QA tests when the small aperture score (with small defined as less than 10 mm) is less than 0.2. Examination of previously published results (Crowe *et al* 2014) indicates that a threshold SAS value of 0.2 also identifies accurate and deliverable prostate IMRT beams (with 100% specificity) from treatments planned with the Brainlab iPlan treatment planning system (using a pencil beam algorithm) for delivery using a Brainlab m3 microMLC. These results suggest that the SAS (and the 0.2 threshold value, for small apertures defined as less than 10 mm) may be used to provide treatment planning guidance at centres using systems other than Varian Eclipse (with AAA) and the MIllennium MLC.

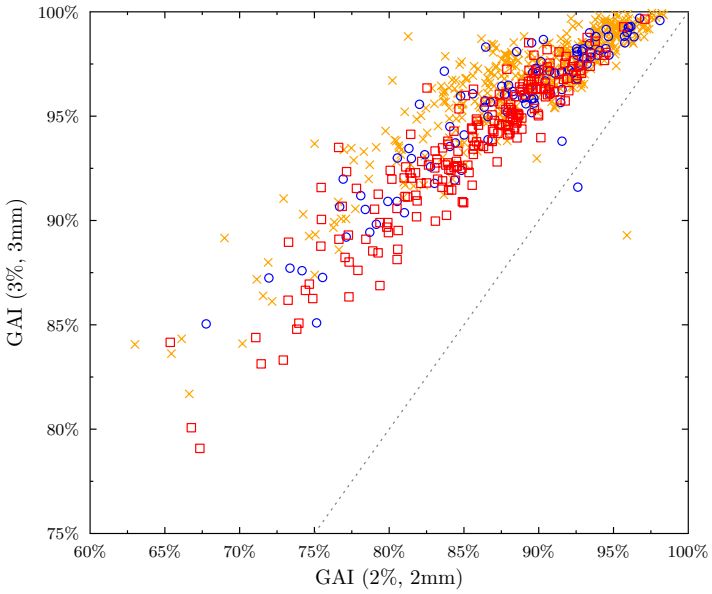

Figure 6. Relationship between GAI calculated using acceptance criteria of 3%, 3 mm and GAI calculated using acceptance criteria of 2%, 2 mm for IMRT beams (orange crosses), VMAT sub arcs (red filled circles) and whole VMAT arcs (blue open circles). Dashed line indicates 1 : 1 relationship. Linear coefficient is 1.085 ± 0.002, with a coefficient of determination of 0.998.

Table 4. Relationship between GAI calculated using acceptance criteria for IMRT and VMAT beams (both whole and sub- arcs).

Treatment	3%, 3 mm Mean GAI (%)	2%, 2 mm Mean GAI (%)	Linear Coefficient	r^2
IMRT	96.48	88.91	1.082 ± 0.002	0.998
Whole arc (360°)	95.34	87.78	1.083 ± 0.005	0.998
Sub-arc (60°)	93.49	85.32	1.094 ± 0.003	0.999
Total	95.44	87.70	1.085 ± 0.002	0.998

Note: r^2 is the coefficient of determination.

By contrast, small aperture scores are generally poorly correlated with the GAI results for VMAT arcs. Comparison of figures 1(*a*)–(*c*), 2(*a*) and 4(*b*) with figures 1(*d*), 2(*b*) and 4(*c*) indicate that the IMRT beams generally produced higher values of metrics correlated with GAI (specifically, the SAS, MCS and MI) than were produced by the VMAT arcs. It is possible that the absence of statistically significant relationships between the GAI values for the VMAT arcs and the metrics evaluated in this work arises from the VMAT arcs being less modulated and using fewer small and asymmetric MLC apertures than were used in the IMRT beams.

4.3. Effects of analysis procedure

The general reduction in the GAI values produced by the VMAT plans, when evaluated as 60° sub-arcs (as opposed to when evaluated as the entire planned 360° arcs), which is illustrated in figure 5, may arise from a genuine increase in treatment delivery inaccuracy due to

the increased proportion of each beam that is involved in starting and stopping, including the ramping up and down of linac output (Kang *et al* 2008), gantry speed and MLC motion. However, the fact that Nicolini *et al* (2008) did not find a similar discrepancy in the evaluation of a small number of sub-arcs (3) with shorter gantry rotations (6° and 12°), despite using similar beam delivery and QA systems, suggest that variations in the performance of the delivery system when delivering arcs of different lengths is unlikely to be the cause of the trend seen in figure 5. Rather, is possible that evaluation of the VMAT plans using sub-arcs reveals disagreements between the planned and delivered doses that are cancelled or averaged out when the entire arc is delivered. This possibility warrants further investigation, since any undetected warping of the dose distribution that could result from increased doses at some angles and decreased doses at others has the potential to result in increased target dose heterogeneity or decreased organ-at-risk sparing.

The linear relationship between GAIs calculated from the quality assurance images using 3%, 3 mm and 2%, 2 mm acceptance criteria, shown in figure 6, has an approximately 8% gradient for both the IMRT and the VMAT beams. Masi *et al* (2013) reported a 7.7% decrease in mean GAI when the gamma evaluation criteria were tightened from 3%, 3 mm to 2%, 2 mm for 142 VMAT treatment plans evaluated using a low-resolution biplanar diode array. These results suggest that a change in gamma acceptance criteria from 3%, 3 mm to 2%, 2 mm is, for most of the beams evaluated in this work, approximately equivalent to tightening the QA passing threshold by 8% . This important result suggests that QA results achieved by different centres using different gamma criteria may be more easily comparable than expected.

5. Conclusions

The large number of treatment plans evaluated in this study allowed an examination of the interrelationships between various new and established modulated beam complexity, deliverability and accuracy metrics and the QA measurement results that many of the metrics were designed to predict.

Variations in metric values were observed for different anatomical treatment sites, suggesting that poorer QA outcomes for anus, endometrium and head and neck treatments arise from the planning system's reliance on more modulated beams consisting of more asymmetric MLC apertures than are needed to cover smaller, less geometrically challenging target volumes.

The IMRT results were able to provide clinically useful guidelines for future treatment planning. In order to produce dosimetrically accurate and clinically deliverable IMRT beams, planners should aim to achieve values of MI less than 0.02, values of MCS greater than 0.4 and CAS less than 0.7. Values of SAS (with MLC apertures narrower than 10 mm defined as 'small') less than 0.2 also identified QA passing IMRT beams with 100% specificity. Plans that fail to meet one or more of these criteria may be automatically replanned or sent for QA testing, depending on local planning and treatment workloads. These criteria also may be broadly applicable to treatment planning and delivery systems other than those used in this study.

Few statistically significant relationships between QA results and beam complexity metrics were identified in the VMAT data, however quarterly review of mean MI, MCS, CAS or SAS values for VMAT plans generated over the preceding three months may prove useful as a means of identifying trends or changes in planning technique or optimiser performance. Additionally, these metrics could prove useful in identifying changes in optimiser performance after any changes or updates to treatment planning software.

Appendix A. List of initialisms

- AAV: aperture area variability
- CAS: cross axis score
- CLS: closed leaf score
- FMC: fluence map complexity
- GAI: gamma agreement index
- MAD: mean aperture displacement
- MCS: modulation complexity score
- MFA: mean field area
- MI: modulation index
- SAS: small aperture score

Acknowledgments

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

References

Ahnesjö A and Aspradakis M M 1999 Dose calculations for external photon beams in radiotherapy *Phys. Med. Biol.* **44** R99

Brainlab AG 2010 Brainlab physics technical reference guide, revision 1.2 *Technical Report* Brainlab AG, Germany

Crowe S B, Kairn T, Kenny J, Knight R T, Hill B, Langton C M and Trapp J V 2014 Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results *Australas. Phys. Eng. Sci. Med.* **37** 475–82

Crowe S B, Kairn T, Middlebrook N, Hill B, Christie D R, Knight R T, Kenny J, Langton C M and Trapp J V 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** 131–8

Ezzell G A *et al* 2009 IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119 *Med. Phys.* **36** 5359–73

Fenoglietto P, Laliberté B, Aillères N, Riou O, Dubois J B and 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

Fogliata A *et al* 2011 Quality assurance of RapidArc in clinical practice using portal dosimetry *Br. J. Radiol.* **84** 535–45

Fredh A, Scherman J B, Fog L S and af Rosenschöld P M 2013 Patient QA systems for rotational radiation therapy: a comparative experimental study with intentional errors *Med. Phys.* **40** 031716

International Commission of Radiation Units and Measurements 2010 *ICRU Report No. 83: Prescribing, Recording, and Reporting Intensity-Modulated Photon-Beam Therapy* (*IMRT*) (Bethesda, MD: International Commission on Radiation Units and Measurements) (doi:10.1093/jicru/ndq001)

Jørgensen M, Hoffmann L, Petersen J B, Praestegaard L, Hansen R and Muren L 2011 Tolerance levels of EPID-based quality control for volumetric modulated arc therapy *Med. Phys.* **38** 1425–34

Kairn T, Crowe S, Kenny J, Knight R and Trapp J 2014 Predicting the likelihood of QA failure using treatment plan accuracy metrics *J. Phys.: Conf. Ser.* **489** 012051

Kairn T, Crowe S, Kenny J and Trapp J 2011a Investigation of stereotactic radiotherapy dose using dosimetry film and Monte Carlo simulations *Radiat. Meas.* **46** 1985–8

Kairn T, Hardcastle N, Kenny J, Meldrum R, Tomé W and Aland T 2011b 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** 333–43

Kang S K, Cheong K H, Hwang T, Cho B C, Kim S S, Kim K J, Oh D H, Bae H and Suh T S 2008 Dosimetric characteristics of linear accelerator photon beams with small monitor unit settings *Med. Phys.* **35** 5172–8

Llacer J, Solberg T D and Promberger C 2001 Comparative behaviour of the dynamically penalized likelihood algorithm in inverse radiation therapy planning *Phys. Med. Biol.* **46** 2637–63

Low D A and Dempsey J F 2003 Evaluation of the gamma dose distribution comparison method *Med. Phys.* **30** 2455–64

Masi L, Doro R, Favuzza V, Cipressi S and Livi L 2013 Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy *Med. Phys.* **40** 071718

McGarry C K, Chinneck C D, OToole M M, OSullivan J M, Prise K M and Hounsell A R 2011 Assessing software upgrades, plan properties and patient geometry using intensity modulated radiation therapy (IMRT) complexity metrics *Med. Phys.* **38** 2027–34

McNiven A L, Sharpe M B and Purdie T G 2010 A new metric for assessing IMRT modulation complexity and plan deliverability *Med. Phys.* **37** 505–15

Miura H, Tanooka M, Fujiwara M, Takada Y, Doi H, Odawara S, Kosaka K, Kamikonya N and Hirota S 2014 Predicting delivery error using a DICOM-RT plan for volumetric modulated arc therapy *Int. J. Med. Phys.: Clin. Eng. Radiat. Oncol.* **3** 82–7

Nicolini G, Fogliata A, Vanetti E, Clivio A, Ammazzalorso F and Cozzi L 2007 What is an acceptably smoothed fluence? Dosimetric and delivery considerations for dynamic sliding window IMRT *Radiat. Oncol.* **2** 42

Nicolini G, Vanetti E, Clivio A, Fogliata A, Korreman S, Bocanek J and Cozzi L 2008 The GLAaS algorithm for portal dosimetry and quality assurance of RapidArc, an intensity modulated rotational therapy *Radiat. Oncol.* **3** 24

Ravkilde T, Keall P J, Grau C, Høyer M and Poulsen P R 2013 Time-resolved dose reconstruction by motion encoding of volumetric modulated arc therapy fields delivered with and without dynamic multi-leaf collimator tracking *Acta Oncol.* **52** 1497–503

Tonigan J, Kry S, Dong L, Purdie T, White R, Ibbott G and Followill D 2011 Does IMRT treatment plan complexity or mismatched dosimetry data contribute to dose delivery errors detected using an IMRT H and N quality assurance phantom? *Med. Phys.* **38** 3804

Tonigan J R 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, Austin

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** 2051–62

Wendling M, Zijp L J, McDermott L N, Smit E J, Sonke J J, Mijnheer B J and van Herk M 2007 A fast algorithm for gamma evaluation in 3d *Med. Phys.* **34** 1647–54