Radiotherapy and Oncology xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original article

Multi-centre audit of VMAT planning and pre-treatment verification

Diego Jurado-Bruggeman [a,*], Victor Hernández [b], Jordi Sáez [c], David Navarro [d], Francisco Pino [e], Tatiana Martínez [f], Maria-Elena Alayrach [g], Norbert Ailleres [h], Alejandro Melero [i], Núria Jornet [j]

[a] Medical Physics and Radiation Protection Department, Institut Català d'Oncologia, Girona; [b] Servei de Física, Hospital Sant Joan, Reus; [c] Servei d'Oncologia Radioteràpica, Hospital Clínic, Barcelona; [d] Servei d'Oncologia Radioteràpica, Consorci Sanitari de Terrassa; [e] Department of Medical Physics, Institut Català d'Oncologia, l'Hospitalet de Llobregat; [f] Department of Radiation Oncology, IMOR Foundation, Barcelona, Spain; [g] Medical Physics Department, Institut Sainte Catherine, Avignon; [h] Département de Cancérologie Radiothérapie, CRLC Val d'Aurelle-Paul Lamarque, Montpellier, France; [i] Department of Radiation Oncology, Institut Català d'Oncologia, Badalona; and [j] Servei de Radiofísica i Radioprotecció, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

ARTICLE INFO

Article history:
Received 12 December 2016
Received in revised form 24 May 2017
Accepted 24 May 2017
Available online xxxx

Keywords:
VMAT
Audit
Quality assurance

ABSTRACT

Background and purpose: We performed a multi-centre intercomparison of VMAT dose planning and pre-treatment verification. The aims were to analyse the dose plans in terms of dosimetric quality and deliverability, and to validate whether in-house pre-treatment verification results agreed with those of an external audit.

Materials and methods: The nine participating centres encompassed different machines, equipment, and methodologies. Two mock cases (prostate and head and neck) were planned using one and two arcs. A plan quality index was defined to compare the plans and different complexity indices were calculated to check their deliverability. We compared gamma index pass rates using the centre's equipment and methodology to those of an external audit (global 3D gamma, absolute dose differences, 10% of maximum dose threshold). Log-file analysis was performed to look for delivery errors.

Results: All centres fulfilled the dosimetric goals but plan quality and delivery complexity were heterogeneous and uncorrelated, depending on the manufacturer and the planner's methodology. Pre-treatment verifications results were within tolerance in all cases for gamma 3%-3 mm evaluation. Nevertheless, differences between the external audit and in-house measurements arose due to different equipment or methodology, especially for 2%-2 mm criteria with differences up to 20%. No correlation was found between complexity indices and verification results amongst centres.

Conclusions: All plans fulfilled dosimetric constraints, but plan quality and complexity did not correlate and were strongly dependent on the planner and the vendor. In-house measurements cannot completely replace external audits for credentialing.

Volumetric-modulated arc therapy (VMAT) has become the most common form of intensity-modulated radiation therapy (IMRT) in most radiotherapy centres and the number of patients being treated is increasing year by year. Pre-treatment verification is recommended for each VMAT plan [1] but treatment unit time for such task is limited.

Faced with this situation, the Catalan Society of Medical Physicists (SCFM) plans to develop recommendations on pre-treatment IMRT verification, and establish criteria to determine which plans should be verified pre-treatment and which plans should not. To meet this objective, four working groups were created. One group developed a software to calculate complexity indices from dicom plans exported from the treatment planning systems, including specific indices for VMAT. Another group developed tools for MLC log-file analysis in order to check MLC positioning deviations and to compare planned fluencies with those obtained from logs [2]. The third group checked the consistency between centres of static gantry IMRT pre-treatment verifications using different measuring equipment and approaches [3]. The fourth group was created to expand the work of the third one to VMAT treatments.

This paper presents the work of the fourth working group. The main aim of this study was to validate the consistency of VMAT pre-treatment verification results in the participating centres by comparing their results with those of an external audit. The results of this study could contribute to the discussion about replacing external dosimetry audits with in-house verifications when credentialing centres for participation in clinical trials [4]. Although it was not the main aim of this group, the planning process was

* Corresponding author at: Medical Physics and Radiation Protection Department, Institut Català d'Oncologia, Avda. França s/n, 17007 Girona, Spain.
E-mail address: djurado@iconcologia.net (D. Jurado-Bruggeman).

http://dx.doi.org/10.1016/j.radonc.2017.05.019
0167-8140/

evaluated by analysing and comparing the dosimetric quality, complexity and delivery of the plans, taking advantage of the software tools developed by the previous working groups.

The audit was promoted within the framework of the Catalan-Occitan Oncology Group to broaden the number of participating centres.

Materials and methods

Nine centres, seven from Catalonia and two from France, participated in this audit. This sample was relatively large and heterogeneous as eleven centres were treating patients with VMAT in the regions being audited, and it comprised public and private institutions, different vendors, TPSs, verification equipment, and methodology.

A form was sent to each centre to collect their technical and dosimetric data, for both planning and pre-treatment verifications. The treatment units and TPSs used by each centre are listed in Table 1. All of them used 6 MV photon beams for VMAT treatments. Eclipse used PRO optimisation algorithm and Monaco used FSPB.

Mock cases

We selected the most clinically relevant tests from those proposed by the American Association of Physicist in Medicine (AAPM) TG119 [5] for IMRT commissioning: test I2 mock prostate and test I3 mock head and neck. We downloaded the structures for these mock cases in DICOM-RT format from the AAPM website and used the dose prescription and optimisation objectives as stated in TG119 (Supplementary Table 1).

For the prostate case, the structures consisted of an ellipsoid (4.0 cm RL, 2.6 cm AP, and 6.5 cm SI) simulating the prostate CTV and PTV, a cylinder with a 1.5 cm diameter simulating the rectum that partially intersected the posterior part of the prostate, and an ellipsoid centred on the top of the prostate simulating the bladder. For the head and neck case, the structures were the PTV, cord, and parotid glands. The PTV was retracted 0.6 cm from the skin. There was a gap of 1.5 cm between the cord and the PTV.

To adapt these mock cases to VMAT, we modelled two cylindrical water-equivalent phantoms that encompassed these structures: one for the prostate case with 25 cm diameter, and another one for the head and neck case with 15 cm diameter. Supplementary Fig. 1 shows both mock cases. Then, we sent the sets of images and structures in DICOM format to all centres.

Treatment planning

All centres planned these mock cases twice, using a single arc and two arcs, and they were asked to report which option they used in clinical practice. The isocentre was placed at the centre of the virtual phantoms, which was inside the PTV for both cases.

All centres were asked to plan these mock cases not only to fulfil the TG119 dose/volume goals, but also to maximise PTV coverage, homogeneity and conformity while minimising the dose to OARs. The plans were compared amongst centres in terms of dose distribution as well as in terms of complexity.

The dosimetric comparison was performed with Sun Nuclear's PlanIQ. For each mock case we defined a Plan Quality Index (PQI), which is a weighted sum of the score of a set of objectives. The scoring depends on the difference between the goal and the achieved value. The definition of the PQI took into account the optimisation goals, reporting considerations from ICRU 83 [6], as well as homogeneity and conformity indices. Planning goals from TG119 and PlanIQ's scoring parameters are shown in Supplementary Table 1. A higher value of PQI means a better dosimetric plan quality as the PTV would be irradiated with the prescribed dose with good conformity and dose homogeneity, and the OAR doses would be minimised. The centres were not aware of the PQI definitions nor of any quantitative criterion for coverage, homogeneity, and conformity evaluation during the planning process to avoid their influence in the process and, therefore, to obtain plans representative of their clinical routine.

The complexity of the plans was analysed using the software developed by the SCFM working group. This software, written in MatLab (MathWorks, Inc.), calculates complexity indices from plans in DICOM format. We chose representative indices for VMAT treatments: the number of MUs and the beam on time needed to deliver the plan [7], the beam irregularity (BI) as defined in [8], the modulation complexity score (MCS) calculated with a minimum gap of 0.5 mm [9], and the total modulation index (MIt) [10]. BI represents the aperture shape irregularity on the basis of beam apertures and MU weights of all segments. MCS represents modulation complexity taking into account the relative variation on leaf positions, beam aperture area, and MU weights between control points. MIt reflects the speed and acceleration of modulating parameters such as MLC movements, dose-rate, and gantry speed. Thus, low complexity plans exhibit BI and MCS values close to 1, and MIt around 0. On the other hand, high complexity plans exhibit high BI and MIt values (no upper limit), and MCS towards 0. Additionally, we assessed the variations of dose-rate and gantry speed between consecutive control points. We also evaluated the fluence as a function of the gantry angle, defined as the product of the beam area and the fraction of MUs delivered at each control point.

Pearson's correlations were used to explore any relationship between PQI and complexity indices.

Pre-treatment verification on-site

All centres were asked to prepare the pre-treatment verification plans as they did in their usual practice. The characteristics of their measuring system, the experimental set-up, and the software and settings for gamma evaluation are shown in Table 2. The measurements were performed together with the external audit.

All centres verified the plans using global gamma evaluation. The gamma pass criterion was of 95% of points with a gamma less

Table 1
Participating centres' equipment.

Centre	Accelerator	MLC	FFF	TPS	Calc. algorithm*	Calc. resolution (mm)	Dos. leaf gap (cm)	Leaf transm. factor
A	Varian Clinac iX	Millennium 120	N	Eclipse v13.0.26	Acuros XB	2.0	0.16	0.014
B	Varian TrueBeam	Millennium 120	N	Eclipse v13.0.26	AAA v10.0.28	2.5	0.14	0.017
C	Varian Clinac iX	Millennium 120	N	Eclipse v13.0.26	AAA	2.5	0.26	0.012
D	Varian Trilogy	HD 120	N	Eclipse v13.5.35	AAA	2.5	0.15	0.012
E	Varian Clinac 2100 C/D	Millennium 120	N	Eclipse v10.0.28	AAA	2.5	0.19	0.016
F	Elekta Synergy	Agility 160	N	Monaco v3.30.01	CVMC	3.0	n/a	0.003
G	Varian Unique	Millennium 120	N	Eclipse v10.0.28	AAA	2.5	0.15	0.015
H	Varian TrueBeam STX	HD 120	N	Eclipse v13.5.35	AAA	2.5	0.12	0.010
I	Varian TrueBeam STX	HD 120	Y	Eclipse v10.0.28	AAA	2.5	0.05	0.012

* Same version as TPS unless specifically mentioned.

Please cite this article in press as: Jurado-Bruggeman D et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol (2017), http://dx.doi.org/10.1016/j.radonc.2017.05.019

Table 2
Pre-treatment equipment and methodology.

	Centre A	Centre B	Centre C	Centre D	Centre E	Centre F	Centre G	Centre H	Centre I
Manufacturer	Sun Nuclear	Sun Nuclear	Sun Nuclear	IBA	IBA	PTW	Sun Nuclear	Varian	Varian
Model	ArcCHECK	ArcCHECK	ArcCHECK	MatriXX Evolution	MatriXX Evolution	Octavius 729	ArcCHECK	aS1200	aS1200
Detectors									
Type	Diode	Diode	Diode	Ion. chamber	Ion. chamber	Ion. chamber	Diode	aSi	aSi
Size (mm)	0.8 × 0.8	0.8 × 0.8	0.8 × 0.8	4.5 diameter	4.5 diameter		0.8 × 0.8	0.34 × 0.34	0.34 × 0.34
Bin (mm)	10	10	10	7.62	7.62	10	10	0	0
Last calibration									
Type	absolute	absolute	relative	absolute	absolute	relative	absolute	absolute	relative
Time until audit	2 months	12 months	40 months	3 months			11 months	4 months	
Set-up									
Phantom	PMMA plug	No	No	Multicube	Multicube	Octavius 4D	No	No	No
Measurement depth (cm)	3.3	3.3	3.3	11	11	16	3.3	0.8	0.8
Predicted dose resolution (mm)	1	1	2.5	1	2.5	2	2.5	0.393	–
Software	Sun Nuclear SNC Patient v6.5.2	Sun Nuclear SNC Patient v6.5.2	Sun Nuclear SNC Patient v6.5.2	IBA OmniPro IMRT v1.7b	IBA OmniPro IMRT v1.7b	PTW Verisoft v6.1	Sun Nuclear SNC Patient v6.2	Varian PDIP[a] v13.3.35	Varian PDIP[a] v10.0.28
Analysis									
Type	Absolute	Absolute	Relative	Absolute	Absolute	Absolute	Absolute	Absolute	Absolute
Threshold	10% max	10% max	10% max	10% max	10% max	10% max	10% max	ROI	–
Output correction	No	No	No	Yes	Yes	Yes	Yes	No	Yes
Interpolation	No	No	No	No	No	Yes	No	–	No
Gamma evaluation									
Arc/total	Total	Total	Total	Total	Total	Total	Total	Arc	Arc
3D gamma[b]	Yes	Yes	No	No	No	Yes	No	No	No

[a] PDIP = Portal Dose Image Prediction.
[b] 3D gamma evaluation uses the TPS 3D dose distribution to find the smallest value of DTA in any direction for each measured point.

than 1 for a 3% dose difference and 3 mm distance to agreement (DTA) in all centres, except centre E which used 2% dose difference and 2 mm DTA. From here on we will refer to this as 3%-3 mm or 2%-2 mm. For this inter-comparison, all centres were asked to submit gamma pass rates for 3%-3 mm and 2%-2 mm criteria.

External audit

For the external audit, a physicist made an on-site visit to each of the participating centres and verified the plans using the same equipment and methodology in all of them. The dosimeters, phantoms, and analysis fulfilled the requirements of the AAPM TG120 [11].

The equipment used for the external audit was ArcCHECK™ with CavityPlug™ from Sun Nuclear Co, and a PTW TM31016 PinPoint 3D (0.016 cc) ionisation chamber. The ArcCHECK™ is made of PMMA and has 1386 diodes arranged with 1 cm spacing in a helical pattern. The CavityPlug™, made of PMMA as well, was fitted in the ArcCHECK™ cavity and the ion chamber was inserted in it, matching the isocenter.

Relative and absolute dose calibrations are required by the ArcCHECK™. The relative calibration was performed in one of the centres before the on-site visits. Its validity amongst the centres was ensured as beam qualities were very homogeneous ($TPR_{20,10}$ between 0.665 and 0.671) and as all measurements were performed within two months. The absolute dose calibration was performed in each centre before the measurements to take into account the particular beam characteristics. The ion chamber and electrometer calibration were traceable to a secondary standard.

To compare the delivered and planned dose distributions, we asked the centres to calculate the dose distribution in the audit equipment. We provided them with a set of images and structures – ArcCHECK™ with CavityPlug™ and the ionisation chamber cavity volume – in DICOM format to import into their TPS. Then, they added their treatment couch and assigned the phantom material properties as specified by the manufacturer for their calculation algorithm. They were asked to calculate the dose distributions with 1 mm resolution, to export them in DICOM-RT format, and to report the doses to the ion chamber in terms of dose-to-water (Dw) – mean and standard deviation.

We also requested the centres to calculate the dose at the ion chamber inside the ArcCHECK™, at 13.3 cm depth, for a 10 × 10 cm^2 field at 0° gantry angle. The measured to calculated dose ratio was used to compensate the chamber readings for daily output variations and to reduce uncertainties in dose calculations in the phantom.

We performed the measurements together with those of the centre. We compared the measured and calculated dose distributions with Sun Nuclear's SNC Patient software (v6.5.2). We calculated global 3D gamma indices for the plans using absolute dose differences, normalised to the maximum calculated dose corresponding to any detector location. The gamma pass rates for 3%–3 mm and for 2%-2 mm with a threshold of 10% of maximum dose were recorded.

We assessed the MLC positional accuracy for all irradiations from Varian machines using MLC log files [12–14]. These log files were analysed with the software developed by the SCFM working group [2] to determine leaf maximum deviation and root mean square (RMS) leaf errors.

Results

Treatment planning

We used the PQI to objectively identify which plans achieved the best trade-off between target coverage, homogeneity, confor-

Please cite this article in press as: Jurado-Bruggeman D et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol (2017), http://dx.doi.org/10.1016/j.radonc.2017.05.019

mity, and doses to OARs. Fig. 1 shows PQI values and the results achieved for each optimisation goal compared to the mean and confidence limits (standard deviation) reported in TG 119. All the centres fulfilled target coverage requirements and dose limits to OAR. For the prostate case, all centres achieved better PTV coverage and lower doses to OARs than the mean of centres participating in TG119 [9], except D10 for the rectum and, for centre F, D10 for the bladder. That was not the case for the head and neck case for which only centres A, B, and I had plans with the same or better fulfilment of all dose goals than those reported in TG119 [9].

The plan quality was better when two arcs were used, especially for the prostate, and only centres B and H get a slightly worse PQI for the head and neck case. In clinical routine, all centres used two arcs for head and neck treatments whereas for prostate a single arc was preferred with the exception of centres A, B, and C.

The relations between plan quality and complexity indices are shown in Table 3. Centre F used the least MUs per plan for prostate whereas for head and neck it used the most. On the other hand, the opposite occurred for centre H. Centre F, the only one using Elekta and Monaco, systematically produced plans with the lowest BI and the highest MCS (factor 2 in both cases), while it used much higher variations of dose rate (factor between 3 and 7), and gantry speed (factor between 3 and 5). Centre D had the highest BI values in all cases and the lowest MCS in head and neck.

We did not find any statistically significant correlation between PQI and complexity. Regarding complexity, there was a correlation between BI and MCS for the whole set of plans ($R = -0.59$, $p < 0.001$). For the prostate case, MUs and BI were correlated ($R = 0.79$, $p < 0.001$), while for the head and neck the correlation was found between MCS and MIt ($R = 0.82$, $p < 0.001$). Beam on time was mainly influenced by the number of arcs, which was expected as gantry speed is limited and the maximum dose rate was barely achieved.

Pre-treatment verification: on-site results and audit results

Fig. 2 shows the gamma pass rates for the on-site pre-treatment verifications and for the audit results. Gamma 3%-3 mm was within 97% in all cases and gamma 2%-2 mm had lower passing rates, where centres D, F and H presented discrepancies above 10% for some plans. On-site results for centres D and H were better than those of the audit while the opposite occurred for centre F.

Point measurements at the isocentre were within ±3% of the calculated doses, except centre D for the head and neck with 1 arc, that exhibited a −8.1% discrepancy. All these measurements were corrected by the difference between calculated and measured doses for the 10×10 cm^2 field, that includes output variations and calculation uncertainties. All output variations were within ±1%. If only the output correction was considered, some differences were found depending on the calculation algorithm. Acuros XB overestimated the dose by 2%, CVMC was below 1% difference, AAA underestimated the dose between 1% and 2%, and for AAA configured at fixed SSD underestimation was about 3% (centres E and I).

Only one correlation was found between complexity indices and audit verification results: MUs and gamma 2%-2 mm ($R = -0.70$, $p < 0.01$).

Leaf positioning errors obtained from log-file analysis for Varian units showed a different behaviour for Clinacs and TrueBeam systems. The positional errors for all Clinacs were very similar irrespective of the plan and the centre, with maximum positional deviations of 1.75 ± 0.28 mm ($k = 1$). All TrueBeam units also behaved similarly, but their maximum positional deviations were

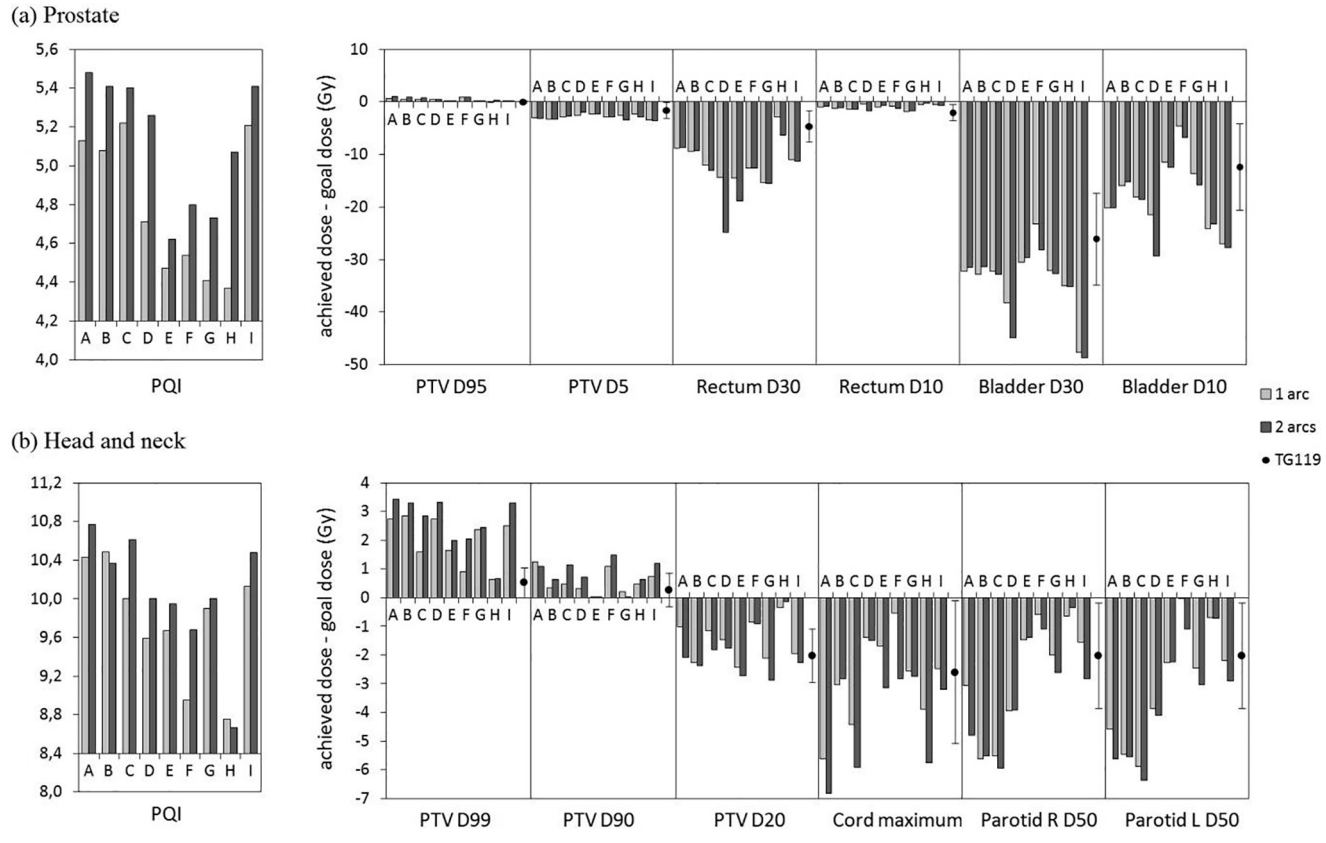

Fig. 1. Plan Quality Index (PQI) values and differences between dose goals and planning results for the prostate case (a) and for the head and neck case (b). Results for the audited centres and the results of TG are compared.

Please cite this article in press as: Jurado-Bruggeman D et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol (2017), http://dx.doi.org/10.1016/j.radonc.2017.05.019

Table 3
Plan quality and complexity parameters.

	No. of arcs	Centre	PQI	MU	Beam on time (s)	BI	MCS	MIt	TotalRRvar (MU/min)	TotalGSvar (deg/s)
Prostate	1	A	5,13	487	74,6	7,6	0,24	0,62	579,6	0,0
		B	5,08	448	74,6	5,7	0,28	0,64	824,4	0,0
		C	5,22	552	74,6	7,3	0,21	0,84	730,8	0,0
		D	4,71	574	78,7	14,9	0,24	0,88	1483,2	7,2
		E	4,47	528	74,6	7,1	0,22	0,66	363,6	0,0
		F	4,54	379	63,4	3,0	0,52	0,87	10461,6	54,0
		G	4,41	545	74,6	7,6	0,24	0,72	759,6	0,0
		H	4,37	684	136,7	11,9	0,24	0,72	0,0	3,6
		I	5,21	560	75,2	10,5	0,26	0,68	1490,4	3,6
Prostate	2	A	5,48	483	149,2	7,3	0,22	0,58	367,2	0,0
		B	5,41	522	149,2	6,7	0,23	0,69	612,0	0,0
		C	5,40	600	149,2	8,7	0,18	0,80	648,0	0,0
		D	5,26	630	149,6	14,7	0,22	0,85	2534,4	0,0
		E	4,62	545	149,2	7,2	0,21	0,61	518,4	0,0
		F	4,80	483	121,3	3,3	0,48	0,76	11066,4	14,4
		G	4,73	561	149,2	7,7	0,23	0,66	813,6	0,0
		H	5,07	654	151,4	13,8	0,24	0,70	705,6	7,2
		I	5,41	508	149,2	10,2	0,27	0,56	2210,4	0,0
Head and neck	1	A	10,43	437	74,6	11,9	0,16	0,83	1281,6	0,0
		B	10,49	451	74,6	10,8	0,20	0,94	1177,2	0,0
		C	10,00	412	74,6	10,8	0,18	0,85	1321,2	0,0
		D	9,59	629	76,7	27,1	0,13	1,05	1461,6	3,6
		E	9,67	482	74,6	11,2	0,18	0,90	2473,2	0,0
		F	8,95	676	107,3	4,9	0,38	1,76	15645,6	108,0
		G	9,90	422	74,6	10,7	0,21	0,90	3348,0	0,0
		H	8,75	373	79,9	14,5	0,26	0,97	338,4	7,2
		I	10,13	412	123,5	22,5	0,22	0,91	0,0	14,4
Head and neck	2	A	10,77	484	149,2	13,8	0,16	0,79	1065,6	0,0
		B	10,37	457	149,2	12,9	0,19	0,87	1360,8	0,0
		C	10,61	440	149,2	12,5	0,16	0,81	1180,8	0,0
		D	10,00	592	149,2	27,4	0,13	0,99	3636,0	0,0
		E	9,95	468	149,2	13,8	0,17	0,82	1814,4	0,0
		F	9,68	647	152,1	4,6	0,36	1,49	19296,0	129,6
		G	10,00	413	149,2	11,4	0,21	0,83	7020,0	0,0
		H	8,67	357	149,2	17,2	0,24	0,84	864,0	0,0
		I	10,48	442	156,1	18,9	0,21	0,90	1987,2	14,4

Plan Quality Index (PQI) and complexity values for each plan: MUs, beam on time, beam irregularity (BI), modulation complexity score (MCS), total modulation index (MIt), total dose rate (TotalRRvar) and gantry speed variations (TotalGSvar). The colours of the heat map grade from the best value – white– to the worst –dark grey– of the parameter for the case considered, irrespective of the number of arcs.

as low as 0.08 ± 0.01 mm (k = 1). The RMS values of leaf positioning deviations are presented in Supplementary Fig. 2, showing again that TrueBeam results were one order of magnitude lower than Clinacs'.

Discussion

To consider the planning and verification processes as a whole could be useful to understand the complex relationship between plan quality, plan complexity, plan deliverability and pre-treatment verification results, especially in a multi-centre environment where multiple planning strategies, TPSs, and linacs from different vendors are involved.

Treatment planning

All centres fulfilled the planning goals, which was not the case for our previous audit on static gantry IMRT [3]. Differences in PQI values between centres were found. As the centres were not aware of the PQI definition and all goals were achieved, it only highlights planning differences and should not be considered to rank them. Differences in complexity indices values between centres were also found, but there was not any correlation between plan quality and complexity. Only BI and MCS indices correlated but MUs and MCS did not, as reported for the multi-institutional audit of VMAT in the UK [15]. Therefore, in a multi-centre environment, a more complex plan delivery is not necessarily associated with a better dose distribution. For the centres using similar technology – machine, TPS, optimisation algorithm, this implies differences in planning strategies regarding dose distribution and delivery. This can be attributable to different planning philosophies and/or planner's expertise. Centres A, B, C, E, G, and I tried to balance dose distribution and deliverability by minimising MUs, gantry speed and dose rate variations, and by checking leaf movements and conformation. Centres D and I prioritised the dose distribution without regard to deliverability as they were confident

Please cite this article in press as: Jurado-Bruggeman D et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol (2017), http://dx.doi.org/10.1016/j.radonc.2017.05.019

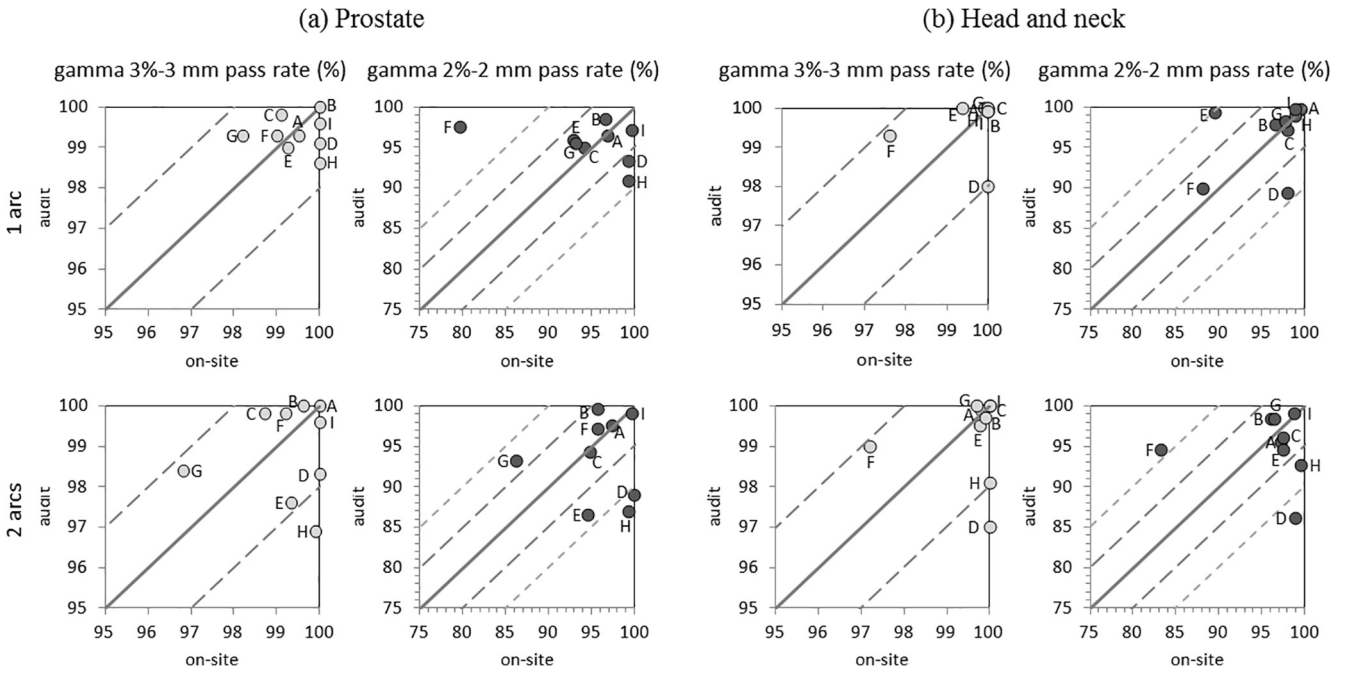

Fig. 2. Gamma pass rates for the on-site pre-treatment verification versus that from the external audit for the prostate and for the head and neck mock cases using one or two arcs. The left and right columns present 3%-3 mm and 2%-2 mm criteria, respectively. The plotted lines delimit the differences between on-site and audit results: solid dark-grey line represents the ideal no difference situation; for 3%-3 mm criteria, the dashed lines mark 2% difference; and for 2%-2 mm criteria, the dark-grey and soft-grey dashed lines mark 5 and 10% differences, respectively.

with their verification methodology. Centre H limited the maximum dose rate to 300 MU/min to facilitate EPID verification, and centre I limited it to 200 MU/min for the head and neck case.

Centre F was the only one not using a Varian machine and Eclipse and its plans were substantially different. Indeed, their PQI values were systematically below the mean and their complexity indices differed from the other centres. In particular, their MUs were the lowest ones for the prostate but the highest ones for the head and neck. BI was much lower and MCS much higher, which means the lowest beam irregularity and modulation; On the other hand, total gantry speed and dose rate variations were much higher than in any other centre, which explains their high MIt values. To investigate the cause of these differences we planned the cases again using Pinnacle^3 Auto-Planning (Philips Radiation Oncology Systems) with a similar machine – manufacturer, model, and MLC. Plan quality improved at the expense of complexity, with PQI values increasing from 4.50 to 5.76 (prostate, 1 arc), from 4.80 to 5.75 (prostate, 2 arcs), from 8.95 to 10.75 (head and neck, 1 arc), and from 9.68 to 10.75 (head and neck, 2 arcs). However, the differences with Varian and Eclipse remained.

These differences can be explained with the help of Fig. 3. It presents gantry speed, dose rate, and fluence at each gantry angle for the prostate case using one arc for four representative centres: two Varians using different machine models, MLC, and TPS versions; centre F; and centre F using Auto-Planning. As can be seen, Varian tended to irradiate continuously from all gantry angles, trying to minimise dose rate changes and, above all, avoiding gantry speed variations that only occurred when dose rate reached its nominal value (see centre I). As a result, MLC shapes and movements were complex and, hence, BI and MCS values were high and low, respectively. On the other hand, Elekta mostly irradiated from specific gantry angles, mimicking static gantry IMRT: gantry rotated at maximum speed at low dose rate until the desired angle was reached and, then, dose rate increased to the maximum value and gantry slowed. Beam shape was regular as the MLC configured a sweeping window that moved fast at these gantry angles. As a result, beam irregularity and modulation were lower, while MIt values were high mainly because of high dose rate and gantry speed variations. All these variations were more pronounced for Pinnacle^3 Auto-Planning plans since the system was pushed to prioritise better dose distributions at the expense of complexity.

Pre-treatment verification on-site versus audit

Our audit results showed that all the participating institutions fulfilled the standards recommended in the Code of Practice for QA and Control for VMAT published by the Netherlands Commission on Radiation Dosimetry (NCS) [16] (3D gamma; 3%-3 mm criterion; pass rate higher than 90%) and in the ESTRO booklet on QA in IMRT (3%-3 mm criterion; pass rate higher than 95%) [1]. Our results are in agreement with those obtained by other groups who reported their results of VMAT audits [17]. However, it is important to remark that the audit and in-house pre-treatment dose verification did not agree for all centres.

The good agreement between in-house and audit results for both 3%-3 mm and 2%-2 mm criteria across centres A, B and C was to be expected because they used the same pre-treatment verification system as the audit. For centre B, in-house results were lower than those of the audit and all of their points failing gamma evaluation were cold points. It was addressed recalibrating the equipment, as the sensitivity of the diodes decreased since the last calibration.

For centre F, using Elekta, the audit results were satisfactory despite the high value of some complexity indices. Nevertheless, in-house results were significantly worse. This was not expected from the sensitivity differences between the ArcCHECK™ and the system used by the centre, Octavius 4D [18,19]. One possible influence could be that Octavius 4D rotates the phantom synchronously with the gantry by means of an inclinometer and a motor. Therefore, pronounced gantry speed changes might desynchronize phantom movement which, in combination with high dose rate

Please cite this article in press as: Jurado-Bruggeman D et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol (2017), http://dx.doi.org/10.1016/j.radonc.2017.05.019

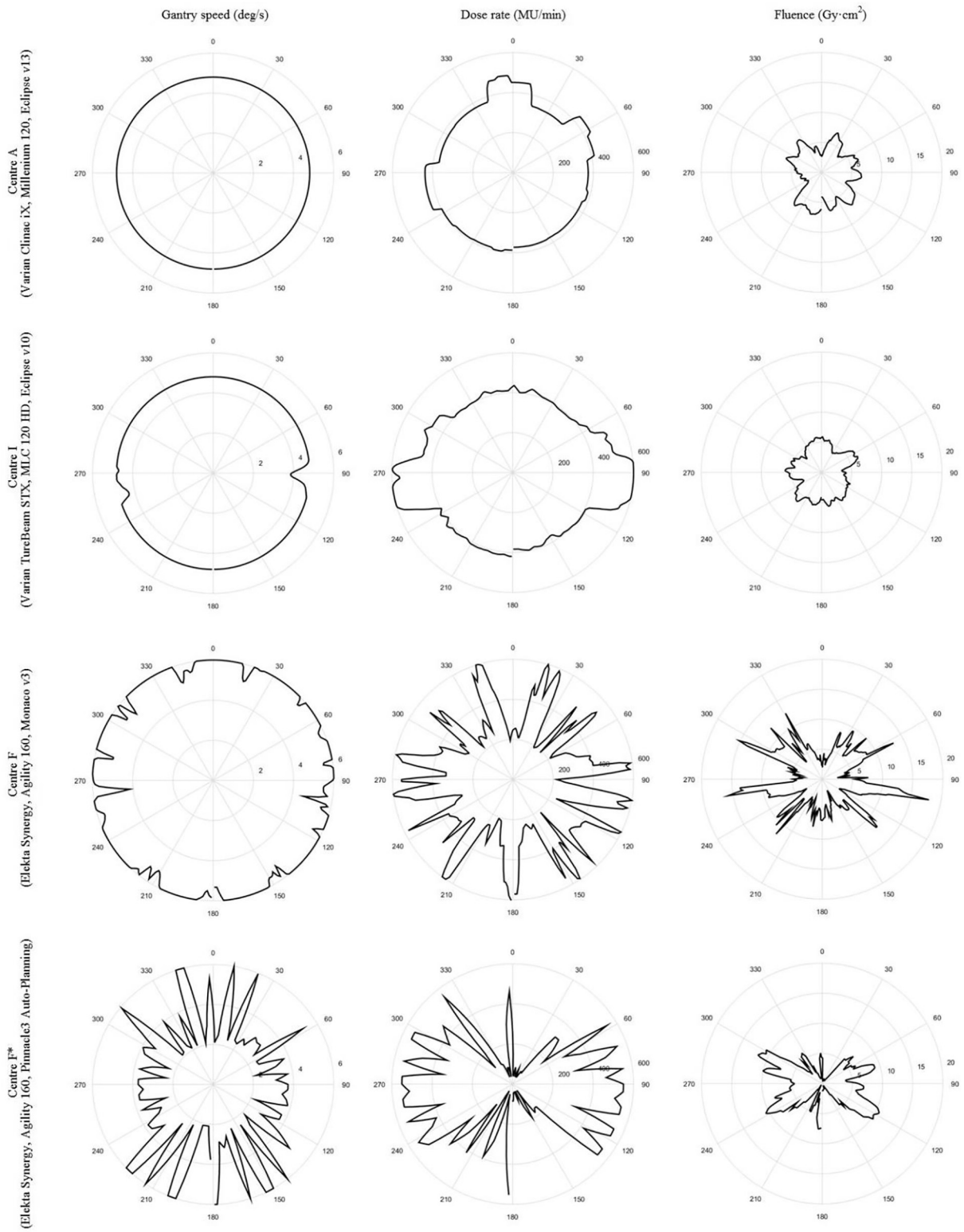

Fig. 3. Gantry speed, dose rate, and fluence as a function of gantry angle for the prostate case using one arc in centres A, I, F, and F using Pinnacle^3 Auto-Planning (F*). The fluence is defined as the product between the field area at the control point and the MUs given in the segment corrected by the Gy/MU at the centre of the ArcCHECK for the 10×10 cm^2 field.

variations at these gantry angles, would negatively affect the results.

For centres using Varian machines, log-file analysis showed that there were not any significant positional errors that could impact the results. There were also no significant differences between in-house and audit deliveries. The results were very similar between linacs of the same type (Clinacs or TrueBeams), and consistent with those of other groups for VMAT [12,14]. These results

Please cite this article in press as: Jurado-Bruggeman D et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol (2017), http://dx.doi.org/10.1016/j.radonc.2017.05.019

are still relevant as they suggest that the differences between in-house and audit verifications should not be attributable to differences in machine delivery and that attention should be diverted to other aspects [12], such as the calculation algorithm, the equipment, or the methodology.

Regarding the calculation algorithm, we were unable to correlate the results of gamma passing rates neither with the algorithm itself nor the software version. This was expected as the pre-measurement absolute dose calibration of the ArcCHECK™ was based on the calculated dose and, thus, encompassed this uncertainty. Nevertheless, some differences arose for the ionisation chamber measurements. The chamber was calibrated in Dw standards, but measurements were performed inside PMMA and, hence, a bias was introduced [20]. Nevertheless, after correcting chamber measurements by the output variation, the relative differences amongst centres can be attributable to the calculation. Calculations were performed modelling PMMA following manufacturer instructions for each algorithm, which were optimised for minimising dose differences at the diodes' depth. Acuros XB and CVMC calculate radiation transport in the medium and, thus, are sensitive to medium composition and characterisation. On the other hand, AAA models the medium as water of different densities. Whether to use one approach or the other, or to report doses in medium or in water requires further work and is out of the scope of this work [21,22]. Although all these differences were within 3%, this should be considered when tightening the acceptance constraints for treatment verification.

In-house gamma passing rates for centres D and H were close to 100% in all cases, although audit results disagreed. Centre D used MatriXX Evolution that had the largest detector size (4.5 mm diameter) which, combined with the high degree of modulation, seems not suitable for 2%-2 mm evaluation. Centre E also used the same system although it was modelled in the TPS as water, and the correction table for oblique beam incidence was the default one. Nevertheless, their modulation was significantly lower and so were the differences. The proportion of cold points for centre D was high and we suspected from problems in daily output measurements. The centre repeated their measurements using its own ArcCHECK and the results were in between those of the MatriXX and the audit, with gamma 2%-2 mm differences with the audit decreasing between 2% and 4% depending on the plan. The discrepancy of 8.1% in the dose at the centre of the ArcCHECK was explained by high dose inhomogeneity caused by MLC complexity (lowest MCS value), that yielded 10% dose variations around 2 mm around this point. Point measurements were sensitive to this dose inhomogeneity, but gamma evaluation using DTA of 2 or 3 mm was not sensitive enough. On the other hand, centre H used EPID images integrated over the whole treatment to perform their verifications. This methodology cannot link any MLC fluence information to gantry angle and, therefore, it should not be used for VMAT verification unless this is taken into account [23,24]. Centre I used the same methodology as centre H, but audit results were satisfactory and consistent with those from the centre. However, they were encouraged to change their methodology as it is potentially insensitive to errors. Octavius 4D, used by centre F, also rotates with the gantry but it does not exhibit this problem because the software uses time-resolved detector readings to obtain 3D dose information.

In general, verification results for two arcs were worse than for a single arc. The complexity of the delivery decreased in most cases as can be seen with MIt, dose rate, and gantry speed variations. However, the overall modulation increased, as can be observed from MCS, MU and BI values, and this circumstance can deteriorate measurement results [16,25]. The decision on whether to use one, two, or more arcs is not straightforward as several factors such as the improvement of the plan, the increase in treatment time, and the influence on QA results should be considered.

No correlation was found between complexity indices and verification results apart from MUs. The use of complexity indices to predict results from pre-treatment verification and to decide their need is, in our opinion, difficult to manage in a heterogeneous multi-centre environment. The involvement of different machines, TPSs, and algorithms, each one with its own particularities and limitations, results in different potential causes of failure for each centre.

From these results, we conclude that using only in-house verifications for VMAT credentialing can be problematic and cannot substitute independent audits. This is in line with recommendations of ESTRO guidelines on IMRT [1], the NSC report on VMAT QA [16], and requirements for participating in clinical trials [4,26,27].

Conclusions

The planning and verification processes were considered globally. Incorporating plan analysis allows identification of differences in treatment delivery parameters that can affect plan complexity and robustness. This analysis was useful to understand the complex relationship between plan quality, plan complexity, plan deliverability and pre-treatment verification results, especially in a multi-centre environment where multiple planning strategies, TPSs, and linacs from different vendors are involved.

We crosschecked the implementation of VMAT in several radiation oncology departments. Regarding plan evaluation and comparisons, all centres fulfilled optimisation goals, improving the results obtained for static gantry IMRT. An index to quantify plan quality is needed for comparison purposes. Gamma 3%-3 mm evaluation requirements were also achieved but differences arose for 2%-2 mm. Aspects as the suitability of the measuring system for VMAT verifications, its characterisation in the TPS, methodology, and agreement on the gamma method used (normalisation, threshold, measuring geometry, etc) are essential to assure reliable results. We cannot recommend using in-house pre-treatment pass rates for multi-centre purposes unless these results have been validated by an external audit.

There was no correlation between plan quality, delivery complexity, and gamma evaluation results amongst centres. This means that, in a multi-centre environment, a more complex plan is not necessarily associated either with a better dose distribution or with better pre-treatment verification results.

Conflict of interest

None.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2017.05.019.

References

[1] ESTRO. Guidelines for the verification of IMRT. 2008. http://www.estro.be/ESTRO/.

[2] Hernandez V, Abella R, Calvo JF, Jurado-Bruggemann D, Sancho I, Carrasco P. Determination of the optimal tolerance for MLC positioning in sliding window and VMAT techniques. Med Phys 2015;42:1911–6. http://dx.doi.org/10.1118/1.4915541.

[3] Jornet N, Carrasco P, Beltrán M, Calvo JF, Escudé L, Hernández V, et al. Multicentre validation of IMRT pre-treatment verification: comparison of in-

house and external audit. Radiother Oncol 2014;112:381–8. http://dx.doi.org/10.1016/j.radonc.2014.06.016.

[4] Weber DC, Poortmans PMP, Hurkmans CW, Aird E, Gulyban A, Fairchild A. Quality assurance for prospective EORTC radiation oncology trials: the challenges of advanced technology in a multicenter international setting. Radiother Oncol 2011;100:150–6. http://dx.doi.org/10.1016/j.radonc.2011.05.073.

[5] Ezzell GA, Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM task group 119. Med Phys 2009;36:5359–73. http://dx.doi.org/10.1118/1.3238104.

[6] Prescribing, recording and reporting photon-beam intensity modulated radiation therapy (IMRT) (ICRU report 83). J ICRU 2010;10:1–35. doi:10.1093/jicru/ndq025

[7] Mohan R, Arnfield M, Tong S, Wu Q, Siebers J. The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy. Med Phys 2000;27:1226. http://dx.doi.org/10.1118/1.599000.

[8] Du W, Cho SH, Zhang X, Hoffman KE, Kudchadker RJ. Quantification of beam complexity in intensity-modulated radiation therapy treatment plans. Med Phys 2014;41:21716. http://dx.doi.org/10.1118/1.4861821.

[9] McNiven AL, Sharpe MB, Purdie TG. A new metric for assessing IMRT modulation complexity and plan deliverability. Med Phys 2010;37:505–15. http://dx.doi.org/10.1118/1.3276775.

[10] Park JM, Park S-Y, Kim H, Kim JH, Carlson J, Ye S-J. Modulation indices for volumetric modulated arc therapy. Phys Med Biol 2014;59:7315–40. http://dx.doi.org/10.1088/0031-9155/59/23/7315.

[11] Low DA, Moran JM, Dempsey JF, Dong L, Oldham M. Dosimetry tools and techniques for IMRT. Med Phys 2011;38:1313–38. http://dx.doi.org/10.1118/1.3514120.

[12] McGarry CK, Agnew CE, Hussein M, Tsang Y, Hounsell AR, Clark CH. The use of log file analysis within VMAT audits. Br J Radiol 2016;89:20150489. http://dx.doi.org/10.1259/bjr.20150489.

[13] Kerns JR, Childress N, Kry SF. A multi-institution evaluation of MLC log files and performance in IMRT delivery. Radiat Oncol 2014;9:176. http://dx.doi.org/10.1186/1748-717X-9-176.

[14] Olasolo-Alonso J, Vázquez-Galiñanes A, Pellejero-Pellejero S, Pérez-Azorín JF. Evaluation of MLC performance in VMAT and dynamic IMRT by log file analysis. Phys Med 2017;33:87–94. http://dx.doi.org/10.1016/j.ejmp.2016.12.013.

[15] McGarry CK, Agnew CE, Hussein M, Tsang Y, McWilliam A, Hounsell AR, et al. The role of complexity metrics in a multi-institutional dosimetry audit of VMAT. Br J Radiol 2016;89:20150445. http://dx.doi.org/10.1259/bjr.20150445.

[16] Mans A, Schuring D, Arends MP, Vugts CAJM, Wolthaus JWH, Lotz HT, et al. The NCS code of practice for the quality assurance and control for volumetric modulated arc therapy. Phys Med Biol 2016;61:7221–35. http://dx.doi.org/10.1088/0031-9155/61/19/7221.

[17] Clark CH, Hussein M, Tsang Y, Thomas R, Wilkinson D, Bass G, et al. A multi-institutional dosimetry audit of rotational intensity-modulated radiotherapy. Radiother Oncol 2014;113:272–8. http://dx.doi.org/10.1016/j.radonc.2014.11.015.

[18] Hussein M, Rowshanfarzad P, Ebert MA, Nisbet A, Clark CH. A comparison of the gamma index analysis in various commercial IMRT/VMAT QA systems. Radiother Oncol 2013;109:370–6. http://dx.doi.org/10.1016/j.radonc.2013.08.048.

[19] Xing A, Arumugam S, Deshpande S, George A, Holloway L, Vial P, et al. SU-E-T-407: evaluation of four commercial dosimetry systems for routine patient-specific tomotherapy delivery quality assurance 319. Med Phys 2014;41. http://dx.doi.org/10.1118/1.4888740.

[20] International Atomic Energy Agency. IAEA technical report series No.398. At Energy 2000:1–229. http://dx.doi.org/10.1097/00004032-200111000-00017.

[21] Ma C-M, Li J. Dose specification for radiation therapy: dose to water or dose to medium? Phys Med Biol 2011;56:3073–89. http://dx.doi.org/10.1088/0031-9155/56/10/012.

[22] Andreo P. Dose to "water-like" media or dose to tissue in MV photons radiotherapy treatment planning: still a matter of debate. Phys Med Biol 2015;60:309–37. http://dx.doi.org/10.1088/0031-9155/60/1/309.

[23] Liu B, Adamson J, Rodrigues A, Zhou F, Yin FF, Wu Q. A novel technique for VMAT QA with EPID in cine mode on a Varian TrueBeam linac. Phys Med Biol 2013;58:6683–700. http://dx.doi.org/10.1088/0031-9155/58/19/6683.

[24] Woodruff HC, Fuangrod T, Rowshanfarzad P, McCurdy BMC, Greer PB. Gantry-angle resolved VMAT pretreatment verification using EPID image prediction. Med Phys 2013;40:81715. http://dx.doi.org/10.1118/1.4816384.

[25] Yang K, Yan D, Tyagi N. Sensitivity analysis of physics and planning SmartArc parameters for single and partial arc VMAT planning. J Appl Clin Med Phys 2012;13:3760. http://dx.doi.org/10.1120/jacmp.v13i6.3760.

[26] Followill DS, Urie M, Galvin JM, Ulin K, Xiao Y, Fitzgerald TJ. Credentialing for participation in clinical trials. Front Oncol 2012;2:198. http://dx.doi.org/10.3389/fonc.2012.00198.

[27] Weber DC, Tomsej M, Melidis C, Hurkmans CW. QA makes a clinical trial stronger: Evidence-based medicine in radiation therapy. Radiother Oncol 2012;105:4–8. http://dx.doi.org/10.1016/j.radonc.2012.08.008.

Please cite this article in press as: Jurado-Bruggeman D et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol (2017), http://dx.doi.org/10.1016/j.radonc.2017.05.019