

Statistical process control to monitor use of a web-based autoplanning tool

Hunter Mehrens^{1,2} | Raphael Douglas¹ | Mary Gronberg^{1,2} | Kelly Nealon^{1,2} |
Joy Zhang¹ | Laurence Court¹

¹Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

²The University of Texas MD Anderson Graduate School of Biomedical Science, Houston, Texas, USA

Correspondence

Hunter Mehrens, Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.
Email: hsmehrens@mdanderson.org

Abstract

Purpose: To investigate the use of statistical process control (SPC) for quality assurance of an integrated web-based autoplanning tool, Radiation Planning Assistant (RPA).

Methods: Automatically generated plans were downloaded and imported into two treatment planning systems (TPSs), RayStation and Eclipse, in which they were recalculated using fixed monitor units. The recalculated plans were then uploaded back to the RPA, and the mean dose differences for each contour between the original RPA and the TPSs plans were calculated. SPC was used to characterize the RPA plans in terms of two comparisons: RayStation TPS versus RPA and Eclipse TPS versus RPA for three anatomical sites, and variations in the machine parameters dosimetric leaf gap (DLG) and multileaf collimator transmission factor (MLC-TF) for two algorithms (Analytical Anisotropic Algorithm [AAA]) and Acuros in the Eclipse TPS. Overall, SPC was used to monitor the process of the RPA, while clinics would still perform their routine patient-specific QA.

Results: For RayStation, the average mean percent dose differences across all contours were $0.65\% \pm 1.05\%$, $-2.09\% \pm 0.56\%$, and $0.28\% \pm 0.98\%$ and average control limit ranges were $1.89\% \pm 1.32\%$, $2.16\% \pm 1.31\%$, and $2.65\% \pm 1.89\%$ for the head and neck, cervix, and chest wall, respectively. In contrast, Eclipse's average mean percent dose differences across all contours were $-0.62\% \pm 0.34\%$, $0.32\% \pm 0.23\%$, and $-0.91\% \pm 0.98\%$, while average control limit ranges were $1.09\% \pm 0.77\%$, $3.69\% \pm 2.67\%$, $2.73\% \pm 1.86\%$, respectively. Averaging all contours and removing outliers, a 0% dose difference corresponded with a DLG value of 0.202 ± 0.019 cm and MLC-TF value of 0.020 ± 0.001 for Acuros and a DLG value of 0.135 ± 0.031 cm and MLC-TF value of 0.015 ± 0.001 for AAA.

Conclusions: Differences in mean dose and control limits between RPA and two separately commissioned TPSs were determined. With varying control limits and means, SPC provides a flexible and useful process quality assurance tool for monitoring a complex automated system such as the RPA.

KEYWORDS

autocontouring, dose verification, statistical process control

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of Applied Clinical Medical Physics* published by Wiley Periodicals, LLC on behalf of The American Association of Physicists in Medicine.

1 | INTRODUCTION

Autocontouring and autoplanning tools are gaining prominence in radiotherapy (RT) research, in part because they may provide better access to RT globally.^{1,2} The Radiation Planning Assistant (RPA) is an integrated web-based autocontouring and autoplanning tool, currently under development, which provides plans calculated on standard linear accelerators (e.g., Varian 2100 using Golden Beam data).³ Users must then recalculate the plans in their own treatment planning systems (TPSs) for their own accelerators. Currently, the RPA provides autocontouring and autoplanning for complex plans, such as volumetric modulated arc therapy (VMAT) of the head and neck^{4–6} and postmastectomy RT to the tangential and supraclavicular fields with field-in-field segments⁷ (called “chest wall” hereafter), as well as simpler plans such as four-field box RT of the uterine cervix^{8–10} and opposed lateral beam treatment of the whole brain.¹¹

However, these techniques also require a rigorous quality assurance (QA) system to ensure safety is maintained and errors are identified within the process of the RPA before treatment delivery.^{1,2} With such a complex system with diverse applications, an intuitive, flexible QA system is needed to monitor plans as they are calculated and exported from different TPSs. In addition, the purpose of this QA application needs to be focused on the process to ensure that communication of changes presented by both the RPA and the user’s TPS do not introduce any additional errors. From a user’s perspective, their standard QA procedure, which may include measurement-based QA, such as gamma analysis for patient-specific QA, has to still be performed outside of the RPA process to inform clinics of the acceptability of plans. Specifically, monitoring dose calculations in the user’s TPS and comparing them with those calculated for standard linear accelerators by the RPA may identify unintentional errors or changes in the overall automated process or in the end user’s systems. For example, changes in dosimetric settings in the end user’s systems, such as the calibration of the planning system or changes to the multileaf collimator (MLC) parameters (e.g., dynamic leaf gap or transmission) may affect dose calculations. Other process changes, intentional or unintentional, that could result in dosimetric changes include changes to the CT-electron density table used for dose calculations and changes to the threshold values used to determine the body contour.¹² Because RPA plans are calculated twice—once by the RPA system and once by the end user’s own TPS—QA systems for RPA should compare the two plans and identify differences or changes over time.

Statistical process control (SPC) has been previously used for dosimetric verification,^{13–17} patient-specific QA,^{18–20} expansion of contour margins,²¹ verification of patient positioning,²² and evaluation of MLC and

machine performance,^{23–25} in image-guided RT²⁶ and adaptive RT.²⁷ The diverse use of SPC provides the flexibility to monitor the RPA’s system. Understanding how SPC means and control limits are affected by differing TPSs and machine parameters may provide insight into how to monitor and assess errors in autocontouring and autoplanning systems. The purpose of this study was to determine the expected differences in means and control limits for different anatomical sites, contours, and machine parameters. Specifically, we used SPC to compare the mean percent dose differences and control limits of different TPS/algorithm combinations for three anatomical sites/treatment approaches (head and neck VMAT, cervix VMAT, and chest wall 3DRT) and machine parameters for head and neck VMAT plans (dosimetric leaf gap [DLG] and MLC transmission factor [MLC-TF]). It is of note that the purpose of monitoring the mean percent dose differences for each contour is to identify gross changes throughout the RPA process.

2 | METHODS

SPC has been used in several applications concerning RT and can help identify gross errors in the overall process. Specifically for our study, here are some examples of errors and their expected impact on the dose distributions that our QA might be able to observe: (a) changes in CT scanner, for example, poor description by the original CT number - electron density tables; (b) changes in MLC parameters, for example, MLC-TF or DLG, which could be changed unintentionally when local TPS is upgraded; and (c) changes in dose calculation algorithm used locally where beam characteristics are not well matched with the original dose calculation algorithm. Several task groups based on TPS commissioning and quality assurance, including TG-53 and TG-157, have shown that beam models can affect accuracy of dose calculations and supposedly, changes in these parameters either intentional or unintentional might produce further errors concerning the dose distribution.^{28,29}

The general workflow of the RPA begins with the upload of a CT scan and a service request form that contains pertinent information such as dose prescription and treatment approach. After the clinical user and radiation oncologist accept the prescription and approach, the RPA automatically performs contouring and planning. The RPA-generated plan is then downloaded from the RPA website and imported into the user’s TPS, in which the user recalculates the dose and completes any desired edits to the plan. Once finalized, the user-calculated plan is uploaded back to the RPA for comparison with the original RPA plan. This workflow was designed to facilitate the deployment of the RPA to many centers without the need to specifically commission the RPA for local linear accelerators, thereby

reducing overall running costs.³ However, the users still need their own TPS.

For this study, RPA plans, including the generated contours, were downloaded and imported into either RayStation v10B (RaySearch Laboratories) or Eclipse v13.6 (Varian) TPS (or both), mimicking an expected clinical workflow. The TPS was then used to recalculate the plans using fixed monitor units within contours provided by the RPA. Following the general workflow, the recalculated plans were uploaded back to the RPA, and the percent mean dose difference between the original RPA plan and the recalculated plan for each contour was calculated.

SPC was used to monitor the variations in percent mean dose difference for each contour. Upper control limits (UCL), lower control limits (LCL), and individual control charts were generated according to Equations (1) and (2)³⁰:

$$\text{UCL} = \mu + 3 * \sigma \quad (1)$$

$$\text{LCL} = \mu - 3 * \sigma \quad (2)$$

where μ is the mean and σ is the standard deviation. Mean percent dose differences that fell outside of the calculated control limits were removed from the calculation of the mean and standard deviation.

The following RT plans were generated using the RPA (Varian Clinac): VMAT for head and neck cancers,^{5,6} VMAT for cervical cancers,^{9,10} and chest wall tangential and supraclavicular fields (with field-in-field segments).⁷ The previous references contain an in-depth view and analysis of how these plans and contours are generated by the RPA.^{5–10} Using the general workflow of the RPA previously mentioned, SPC was used to compare RPA through two methods: (i) anatomical and TPS/algorithmic differences, and (ii) machine parameter differences. It is important to note that the comparison by TPS/algorithm used separate systems, commissioned separately by different teams. This limits the true direct comparison of the TPS/algorithms; however, the flexibility of SPC, specifically its monitoring of trends over time and the process of the system, allows us to monitor changes based on their differences to identify potential errors or changes to the user's independent system. In addition, the anatomical comparison between metrics for the target and structures was based solely on anatomical regions and their corresponding structures that are supported by the RPA, no structural delineation comparison was performed.

2.1 | Anatomical and TPS/algorithmic differences in control limits

To determine how control limits varied by the anatomical site and by differences in TPS and dose calculation algorithm, we generated 32 head and neck VMAT plans, 33

cervical four-field box RT plans, and 51 chest wall plans with the RPA using an Analytical Anisotropic Algorithm (AAA) (Eclipse v.13.6). Specifically, the head and neck VMAT plans consisted of three arcs with energy of 6 MV, cervical four-field box plans had an energy of 18 MV, and chest wall plans had 12 beams at energy of 6 MV and 18 MV with an additional two supraclavicular fields with energy of 18 MV. These plans were downloaded and imported into the TPSs. The dose calculation algorithm used by the RayStation v10B TPS was collapsed cone convolution (CCC), and the Eclipse v13.6 dose calculation algorithm was Acuros (dose to medium). The resulting plans were compared to the original RPA plans, and for each contour, a percent mean dose difference was calculated. Using Equations (1) and (2), control limits were calculated on the basis of these percent mean dose differences.

2.2 | Machine parameter differences in control limits

To determine how control limits varied by machine parameter and dose calculation algorithm in the same TPS, we generated 30 head and neck VMAT plans using the RPA, downloaded them, and imported them into Eclipse v13.6. Dose calculation was performed on a different machine from that used to calculate the RPA plans and used either AAA or Acuros algorithm. In addition, DLG and MLC-TF were varied for each plan. The choice of DLG and MLC-TF was based on a study by Glenn et al.³¹ that found that for the Eclipse TPS, these parameters had the greatest impact on dose changes among treatment parameters. Using the 2.5th–97.5th percentiles of the values determined by Glenn et al., we selected five values for DLG (0.1, 0.155, 0.17, 0.19, 0.23 cm) and four values for MLC-TF (0.0118, 0.0145, 0.0158, 0.0165).^{31,32} The DLG and MLC-TF values for RPA plans were 0.2 cm and 0.02, respectively. While the value of either DLG or MLC-TF was varied, the other parameter value was held constant at that of the RPA plan. The resulting plans with variation in DLG and MLC-TF values were compared to the original RPA plans (i.e., Eclipse-AAA), and for each contour, a percent mean dose difference was calculated. Using Equations (1) and (2), control limits were calculated based on these percent mean dose differences. Linear regression was performed to characterize how the control limits and means changed with variations in DLG and MLC-TF.

3 | RESULTS

3.1 | Anatomical and TPS/algorithmic differences in control limits

Table 1 provides the UCL, mean percent difference, and LCL for a selection of contours for each anatomical site.

TABLE 1 Percent mean dose differences from the RPA plan for selected contours from three anatomical sites

Site/contour	RayStation-CCC mean (LCL, UCL) ^a	Eclipse-Acuros mean (LCL, UCL) ^a
Head and neck		
Brain	-0.21 (-0.44, 0.01)	-0.24 (-0.42, -0.05)
CTV1	2.19 (1.04, 3.34)	-0.25 (-1.25, 0.75)
CTV2	1.26 (0.11, 2.40)	-0.22 (-0.72, 0.28)
CTV3	1.11 (-0.09, 2.31)	-0.36 (-0.82, 0.10)
Oral cavity	1.47 (0.24, 2.71)	-0.37 (-0.72, -0.02)
Left submandibular gland	1.71 (0.14, 3.27)	-0.94 (-1.59, -0.28)
Right submandibular gland	1.74 (0.45, 3.04)	-0.98 (-1.57, -0.39)
Larynx	2.28 (1.09, 3.47)	-0.97 (-1.57, -0.38)
Mandible	1.11 (0.20, 2.02)	-1.60 (-2.39, -0.81)
PTV1	2.29 (1.00, 3.59)	-0.48 (-1.83, 0.88)
PTV2	2.18 (-1.19, 5.55)	-0.15 (-0.68, 0.37)
PTV3	1.55 (-0.03, 3.14)	-0.32 (-0.74, 0.10)
Left parotid	0.49 (-0.61, 1.60)	-0.95 (-1.33, -0.57)
Right parotid	0.43 (-0.63, 1.49)	-0.93 (-1.20, -0.66)
Cervix		
Left femoral head	-1.78 (-2.78, -0.77)	0.21 (-0.92, 1.34)
Right femoral head	-1.79 (-2.81, -0.77)	0.19 (-0.93, 1.31)
L4	-2.21 (-4.87, 0.45)	0.49 (-4.43, 5.41)
L5	-2.59 (-3.43, -1.74)	0.48 (-1.31, 2.27)
Chest wall		
Clinical chest wall	1.51 (-1.65, 4.67)	-1.55 (-3.62, 0.52)
Clavicle	-0.86 (-2.20, 0.49)	-3.60 (-5.64, -1.56)
Heart	-0.17 (-0.68, 0.35)	-0.10 (-0.68, 0.49)
Humeral head	-0.34 (-1.18, 0.50)	-0.52 (-1.01, -0.03)
Left lung	-0.40 (-1.66, 0.86)	-0.36 (-1.97, 1.25)
Right lung	-0.34 (-1.69, 1.02)	-0.32 (-1.79, 1.15)
Ribcage	-0.17 (-1.05, 0.71)	-1.99 (3.59, -0.39)
Supraclavical	0.23 (-1.33, 1.78)	-1.21 (-2.69, 0.28)
Spinal canal	-0.17 (-0.51, 0.17)	-0.14 (-0.35, 0.07)
Sternum	0.61 (-0.56, 1.78)	-0.48 (-1.53, 0.57)

Abbreviations: CCC, collapsed cone convolution; CTV, clinical target volume; LCL, lower control limit; PTV, planning target volume; RPA, Radiation Planning Assistant; UCL, upper control limit.

^aAll values are percentages.

Specifically, head and neck and chest wall cases showed approximately 70% of contours being less than 2% in the mean differences between separately commissioned planning systems (RayStation-CCC and Eclipse-Acuros). All percent dose differences are relative to the automated plan.

The individual control charts for head and neck contours showed that mean percent dose differences and control limits varied by structure and by the TPS/algorithm combination (Figure 1). In general, the differences for organs at risk (OARs) were smaller than those for targets. For RayStation-CCC, the aver-

age range of the control limits (UCL-LCL) was $1.89\% \pm 1.32\%$, with an average mean percent dose difference across all contours of $0.65\% \pm 1.05\%$. In contrast, Eclipse-Acuros' average range was smaller, at $1.09\% \pm 0.77\%$, with an average mean percent dose difference across all contours of $-0.62\% \pm 0.34\%$. Figure 1 shows example contours where this can be identified, especially in the primary target. Comparison of the two TPS/algorithm combinations showed that the average absolute difference in means was $1.31\% \pm 1.05\%$ and the average absolute difference in range was $0.88\% \pm 1.12\%$. Overall, RayStation-CCC

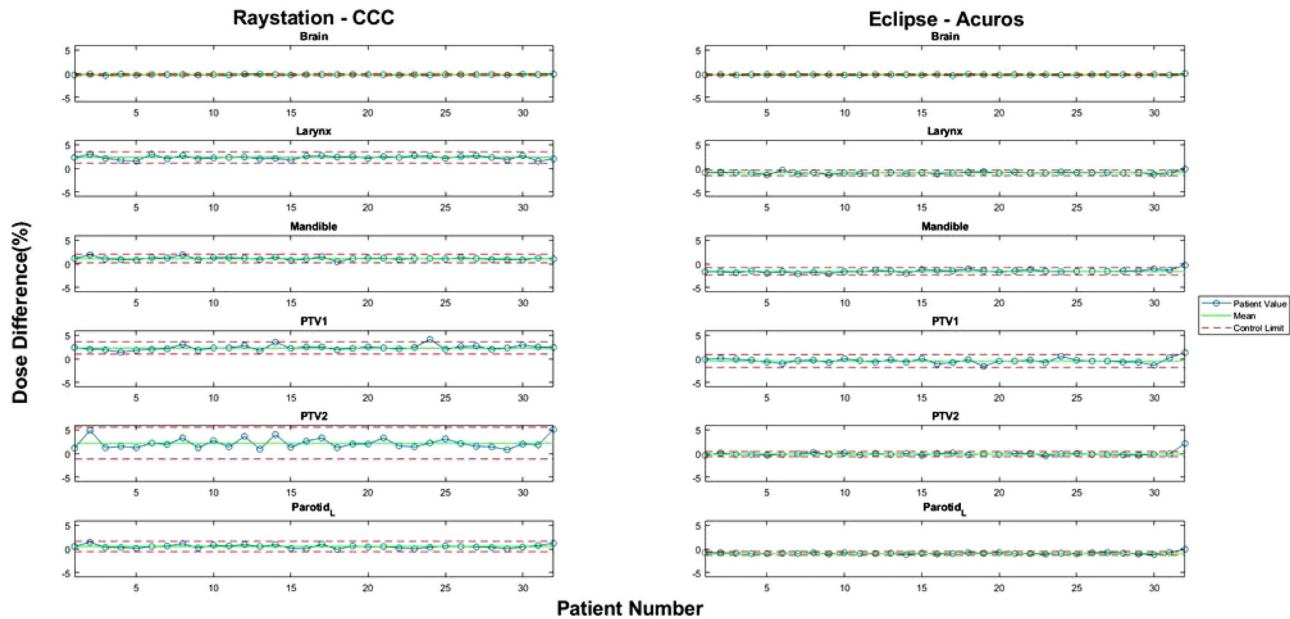


FIGURE 1 Selected individual control charts for head and neck contours. Left: percent dose difference between RPA and RayStation-CCC. Right: percent dose difference between RPA and Eclipse-Acuros. Blue circles: individual cases. Solid green lines: mean percent dose difference calculated by SPC for the indicated contours. Dotted red lines: upper and lower control limits calculated by SPC for the same contours. RPA, Radiation Planning Assistant; CCC, collapsed cone convolution; SPC, statistical process control; PTV, planning target volume

produced a higher dose than did the RPA plan, whereas Eclipse-Acuros produced a smaller dose than the RPA plan.

Unlike the head and neck contours, the mean percent dose differences for cervix contours showed larger and more consistent discrepancies between RPA and RayStation-CCC as opposed to Eclipse-Acuros plans (Figure 2a). The average mean percent dose difference for all cervix contours was $-2.09\% \pm 0.56\%$ for RayStation-CCC, but only $0.32\% \pm 0.23\%$ for Eclipse-Acuros. Nonetheless, the average range of percent dose differences was larger for Eclipse-Acuros, at $3.69\% \pm 2.67\%$, than for RayStation-CCC, at $2.16\% \pm 1.31\%$. The average absolute difference in means across the cervix contours between RayStation-CCC and Eclipse-Acuros was $2.41\% \pm 0.77\%$, with an average absolute difference in range of $1.56\% \pm 1.49\%$. Unlike the head and neck plans, cervix plans with RayStation-CCC produced a lower dose than the RPA plan, whereas Eclipse-Acuros plans had a slightly higher dose than the RPA plan.

Finally, for chest wall plans with tangents and supraclavicular fields, we observed similar differences in means and control limits to those found in head and neck plans (Figure 2b). For RayStation-CCC, the average range of the control limits was $2.65\% \pm 1.89\%$, with a small average mean percent dose difference across all contours of $0.28\% \pm 0.98\%$. In contrast, the average range of Eclipse-Acuros was slightly larger, at $2.73\% \pm 1.86\%$, with a much larger average mean percent dose difference across all con-

tours of $-0.91\% \pm 0.98\%$. The average absolute difference in means across the chest wall contours between RayStation-CCC and Eclipse-Acuros was $1.25\% \pm 1.46\%$, with an average absolute difference in range of $0.83\% \pm 0.81\%$. However, in contrast to both head and neck and cervix plans, RayStation-CCC and Eclipse-Acuros plans both tended to produce a lower dose than the RPA plans. However, several contours, especially those generated by RayStation-CCC, had a higher dose than that of the RPA, which made the overall average positive. As with the head and neck plans, this effect was associated more with targets than with OARs.

3.2 | Machine parameter differences in control limits

Figure 3 illustrates how the control limits and mean percent dose differences for the same six head and neck structures shown in Figure 1 changed along with DLG, and Table 2 provides the rate of change in the control limits and mean of the percent dose differences per 0.1 cm of DLG (and 0.01 of MLC-TF) for Acuros and AAA. These rates of change values for DLG and MLC-TF were chosen so as to discuss the largest variation observed based on what has been reported clinically for varying institutions.³¹ For Acuros, the control limits tended to shrink toward the mean as DLG increased. This feature is especially visible in Figure 3e for the mandible. However, for AAA, the rates of change for UCL,

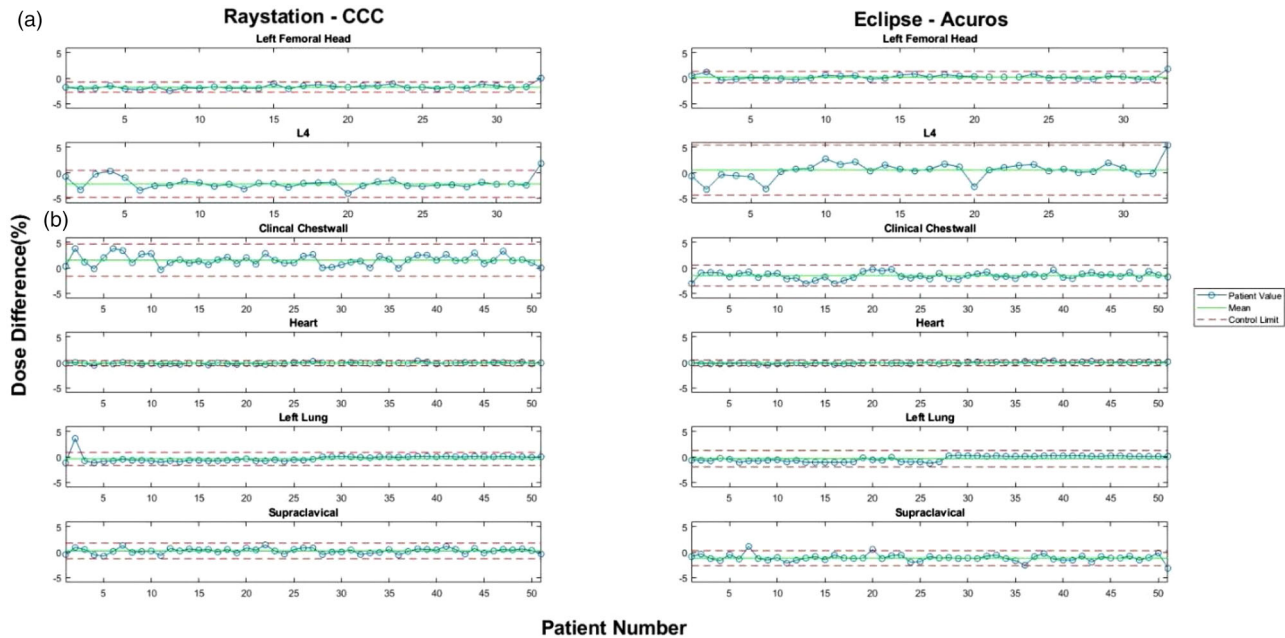


FIGURE 2 Selected individual control charts for (a) cervix and (b) chest wall contours. Left: percent dose difference between RPA and RayStation-CCC. Right: percent dose difference between RPA and Eclipse-Acuros. Blue circles: individual cases. Solid green lines: mean percent dose difference calculated by SPC for the indicated contours. Dotted red lines: upper and lower control limits calculated by SPC for the same contours. RPA, Radiation Planning Assistant; CCC, collapsed cone convolution; SPC, statistical process control

TABLE 2 Rate of change of percent dose differences per 0.1 cm DLG and 0.01 MLC-TF for selected contours of head and neck plans calculated by Acuros and AAA

	Rate of change of percent dose differences/0.1 cm of DLG						Rate of change of percent dose differences/0.01 of MLC-TF					
	Acuros			AAA			Acuros			AAA		
	UCL	Mean	LCL	UCL	Mean	LCL	UCL	Mean	LCL	UCL	Mean	LCL
Brain	0.16%	0.19%	0.21%	0.32%	0.20%	0.09%	0.26%	0.42%	0.57%	0.44%	0.44%	0.44%
Larynx	2.12%	2.45%	2.78%	2.28%	2.54%	2.80%	2.46%	2.46%	2.37%	2.72%	2.49%	2.26%
Mandible	1.01%	1.87%	2.73%	1.78%	2.02%	2.26%	2.32%	2.28%	2.23%	2.47%	2.36%	2.26%
PTV1	2.88%	3.17%	3.45%	3.41%	3.23%	3.05%	1.87%	1.89%	1.91%	2.13%	1.93%	1.74%
PTV2	1.94%	2.25%	2.56%	1.90%	2.28%	2.65%	1.88%	1.85%	1.82%	1.95%	1.87%	1.83%
Left parotid	1.51%	1.74%	1.96%	1.72%	1.81%	1.89%	2.52%	2.54%	2.55%	2.59%	2.59%	2.58%

Abbreviations: AAA, Analytical Anisotropic Algorithm; DLG, dosimetric leaf gap; LCL, lower control limit; MLC-TF, multileaf collimator transmission factor; PTV, planning target volume; UCL, upper control limit.

mean, and LCL were similar for five of the six structures, indicating that the control limits were relatively independent of DLG. However, there were exceptions, for example, the brain (Figure 3a), which showed a widening of control limits as DLG increased. On average for all contours, a 0% dose difference corresponded with a DLG value of 0.299 ± 0.180 cm for Acuros and 0.135 ± 0.031 cm for AAA. However, the Acuros value was greatly skewed by several outliers, including values greater than 0.5 cm for structures, such as the eye, lens, and optic nerve. After we removed these outliers, a 0% dose difference for Acuros corresponded with a DLG value of 0.202 ± 0.019 cm. The overall changes in

control limits and means for AAA and Acuros for all contours for two DLG values (0.1 and 0.19 cm) can be found in Figure 4. A comparison of Figures 4a and 4b shows that targets' means and control limits changed more dramatically as the DLG value increases. In contrast, the means and control limits of OAR structures tended to not change as dramatically.

For the same six head and neck structures, Figure 5 shows how changes in MLC-TF affected mean dose differences. Unlike DLG, Acuros tended to show no consistent distinguishing pattern as MLC-TF increased. This is especially apparent in a comparison of Figures 5a and 5b, which shows a decrease of control limits and no

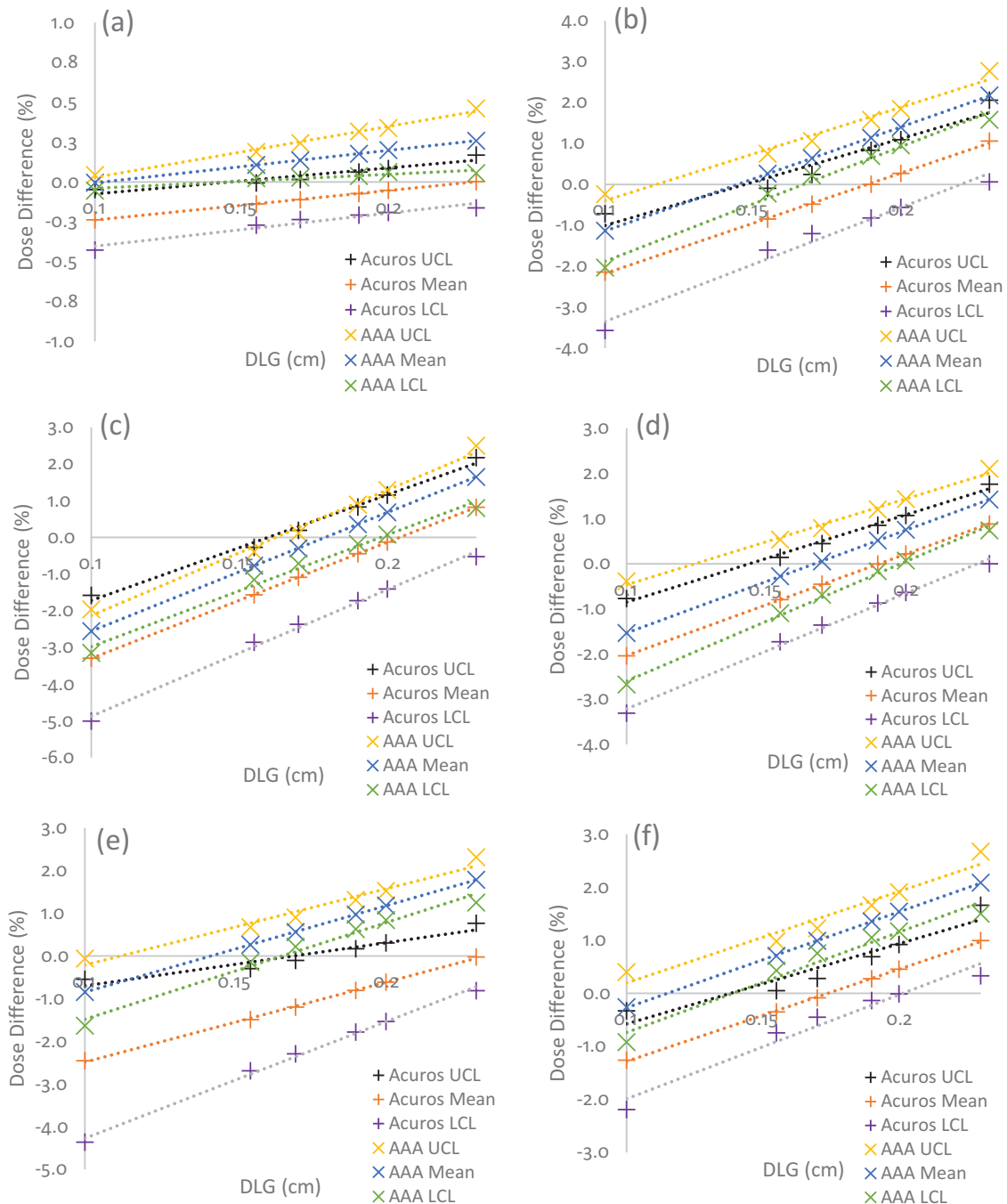


FIGURE 3 Percent dose differences with changes in DLG for head and neck plans calculated with different dose calculation algorithms: (a) brain, (b) larynx, (c) PTV1, (d) PTV2, (e) mandible, and (f) left parotid. The black, orange, and purple + symbols represent changes in Acuros' UCL, mean, and LCL, respectively. Yellow, blue, and green x symbols represent changes in AAA's UCL, mean, and LCL, respectively. The dotted lines represent the linear fit. AAA had a narrower range between control limits than Acuros. For most contours, Acuros matched the RPA plan at an approximate DLG of 0.2 cm (removing outliers), while AAA matched the RPA plan at an approximate DLG of 0.14 cm. DLG, dosimetric leaf gap; PTV, planning target volume; UCL, upper control limit; LCL, lower control limit; AAA, Analytical Anisotropic Algorithm; RPA, Radiation Planning Assistant

change in control limits as MLC-TF increased, respectively. A similar pattern was observed for AAA. Only the control limits for planning target volume 1 (PTV1) showed a slight broadening with increasing MLC-TF, while the control limits for the other contours remained

steady, indicating, as with DLG, that the control limits were relatively independent of MLC-TF. On average, for all contours, a 0% dose difference corresponded with an MLC-TF value of 0.025 ± 0.010 for Acuros and 0.015 ± 0.001 for AAA. Again, we observed that

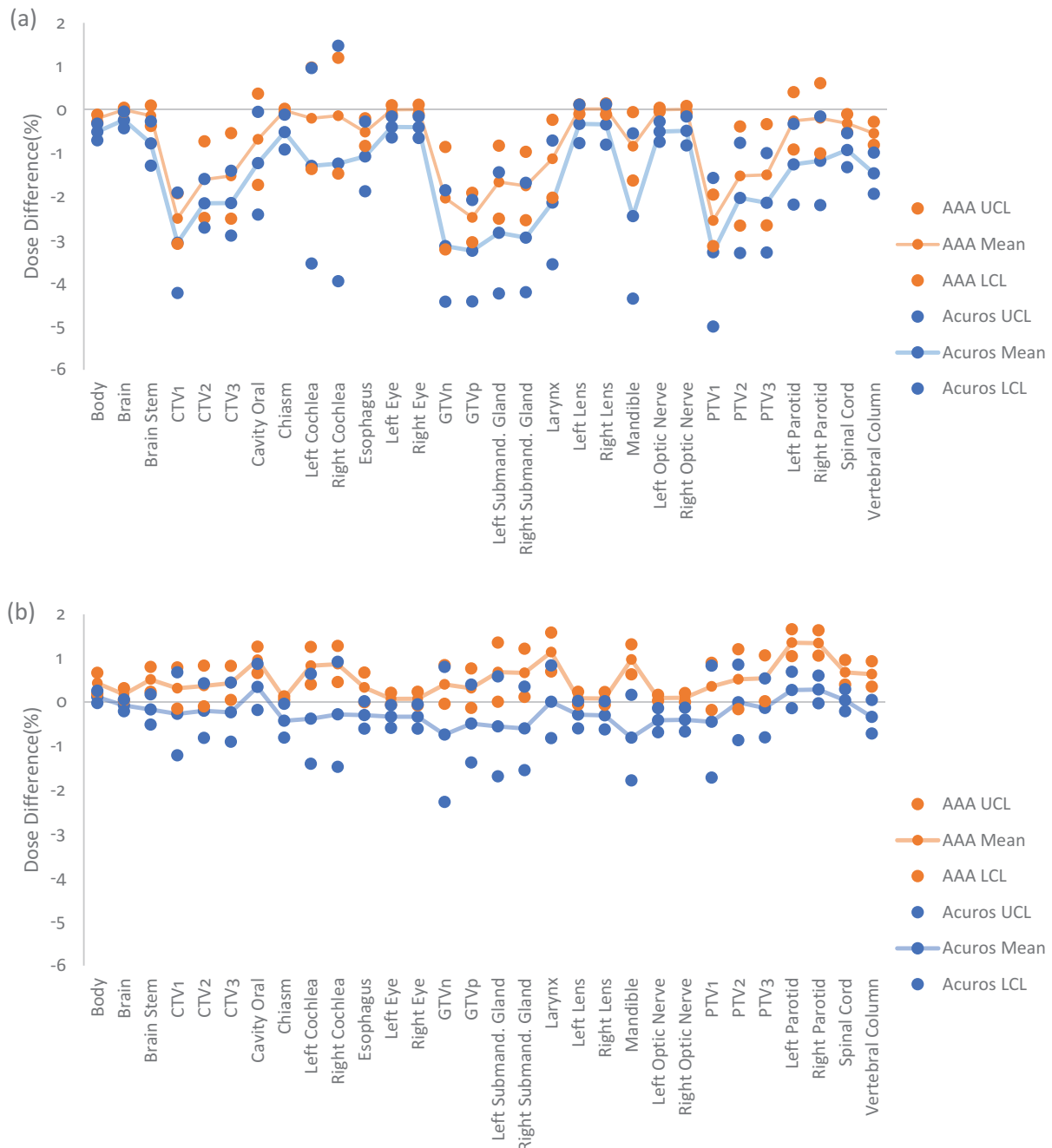


FIGURE 4 Mean percent dose differences, UCL, and LCL for head and neck contours calculated with AAA (orange) and Acuros (blue) at a DLG of (a) 0.1 cm, and (b) 0.19 cm. As DLG increased toward the value of the optimized DLG (0.2 cm) for the RPA machine, Acuros' and AAA's means and control limits became more similar for all contours. UCL, upper control limit; LCL, lower control limit; AAA, Analytical Anisotropic Algorithm; DLG, dosimetric leaf gap; RPA, Radiation Planning Assistant; CTV, clinical target volume; GTVn, gross nodal tumor volume; GTVp, gross primary tumor volume; PTV, planning target volume

this value for Acuros was skewed by several outliers for structures such as the eye, lens, and optic nerves, although the effect of these outliers was not as prominent as it was for DLG. After removal of the outliers, a 0% dose difference for Acuros corresponded to an MLC-TF value of 0.020 ± 0.001 . The overall changes in control

limits and means for AAA and Acuros for all contours for two MLC-TF values (0.0118 and 0.0165) are shown in Figure 6. However, in contrast with the pattern observed for DLG, we observed a consistent rise in the means for all contours as MLC-TF increased, with the control limits at approximately the same range around the mean.

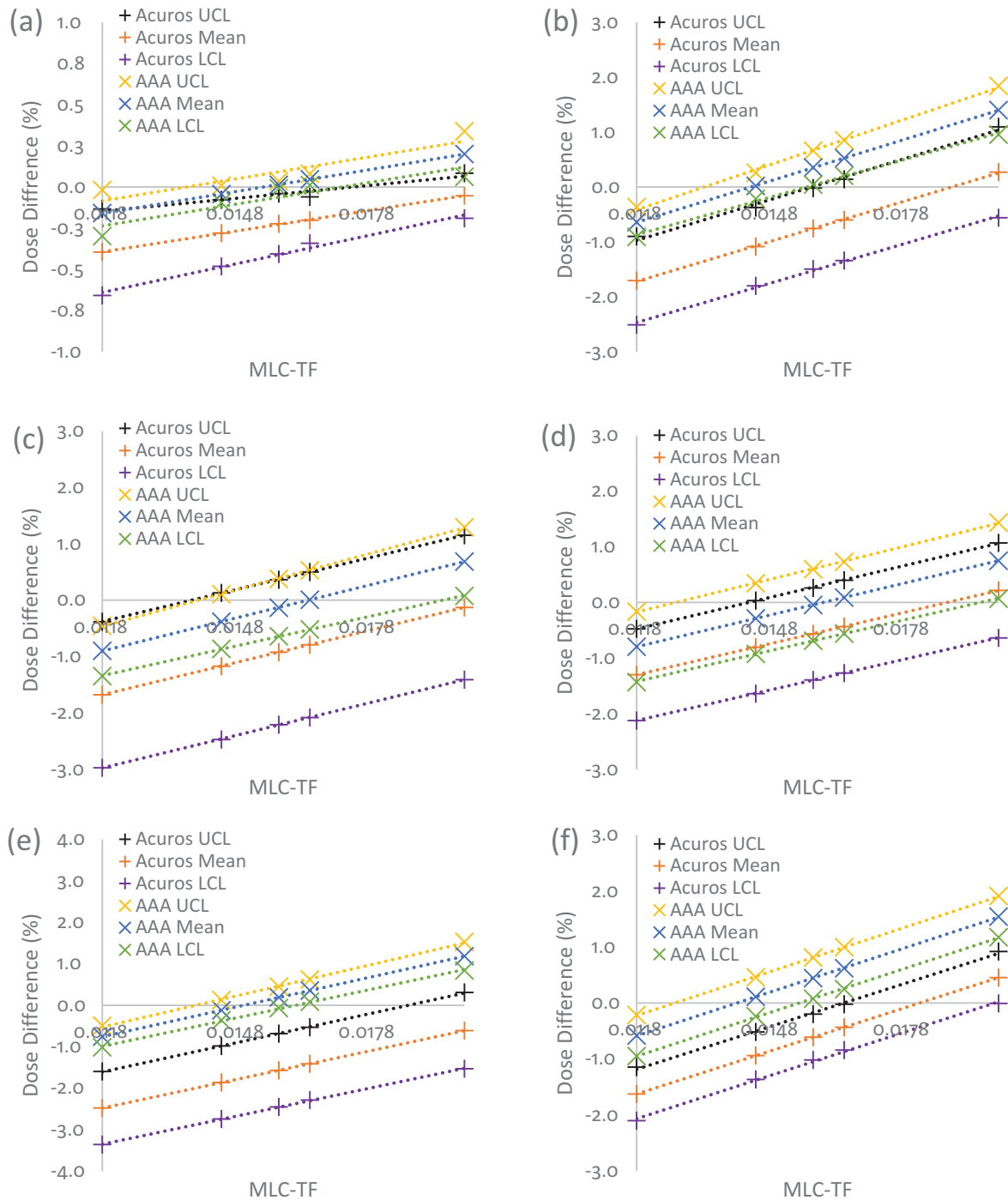


FIGURE 5 Percent dose differences with changes in MLC-TF for head and neck plans calculated for different dose calculation algorithms: (a) brain, (b) larynx, (c) PTV1, (d) PTV2, (e) mandible, and (f) left parotid. The black, orange, and purple + symbols represent changes in Acuros' UCL, mean, and LCL, respectively. Yellow, blue, and green x symbols represent changes in AAA's UCL, mean, and LCL, respectively. The dotted lines represent the linear fit. AAA had a narrower range between control limits than Acuros. For most contours, Acuros matched the RPA plan approximately at an MLC-TF of 0.02, while AAA matched the RPA plan approximately at an MLC-TF of 0.015. MLC-TF, multileaf collimator transmission factor; PTV, planning target volume; UCL, upper control limit; LCL, lower control limit; AAA, Analytical Anisotropic Algorithm; RPA, Radiation Planning Assistant

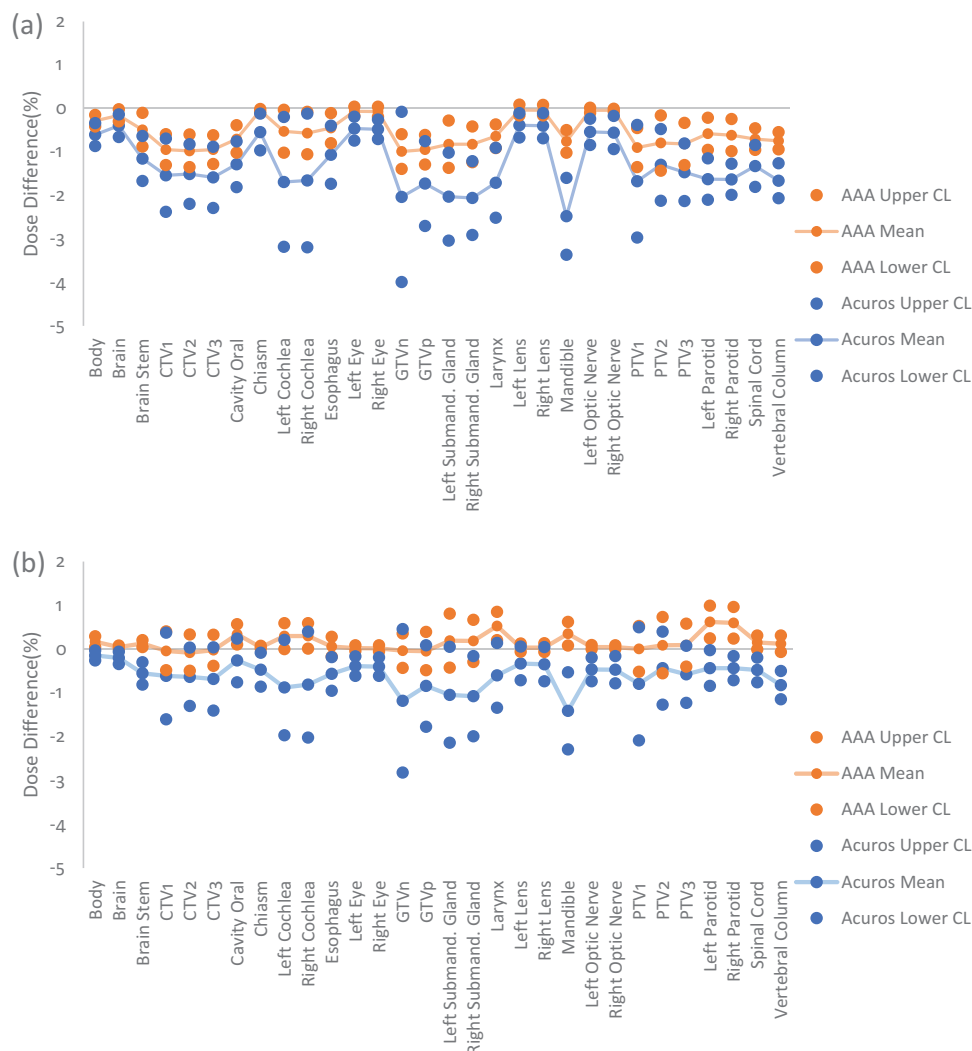


FIGURE 6 Mean percent dose differences, UCL, and LCL for head and neck contours calculated with AAA (orange) and Acuros (blue) with an MLC-TF of (a) 0.0118, and (b) 0.0165. As MLC-TF increased toward the value of the optimized MLC-TF (0.020) for the RPA machine, Acuros' and AAA's means and control limits became more similar for all contours; however, this effect was less pronounced than it was for DLG. UCL, upper control limit; LCL, lower control limit; AAA, Analytical Anisotropic Algorithm; MLC-TF, multileaf collimator transmission factor; RPA, Radiation Planning Assistant; DLG, dosimetric leaf gap; CTV, clinical target volume; GTVn, gross nodal tumor volume; GTVp, gross primary tumor volume; PTV, planning target volume

4 | DISCUSSION

In general, several factors must be considered when using SPC to monitor a process. These can include, but are not limited to, the known potential systematic errors throughout a process and the effect of the process on the metric you are monitoring. Furthermore, the implementation of SPC is important considering which control charts might be used (e.g., individual control charts) and how to identify the number of points that are sufficient to categorize SPC's mean and standard deviation. However, for our specific case, SPC will be integrated into the RPA as a process QA monitoring system and the determination of these important considerations concerning SPC will be determined and adapted by the RPA staff. We aim to flag outlier situations to the user, rather than have them develop their own SPC. For further reference

and a graphical representation of the RPA workflow, please see the several figures provided by Court et al.³

SPC allows for the monitoring of these systems over time and can alert the user or provider of these tools to systemic changes or errors that may affect plans. However, although a fixed threshold (e.g., 5%) may be sufficient to identify large errors, this approach may not be sufficient to capture the numerous potential errors that may arise during the complex process of auto-contouring and autoplanning. Failure to identify small potential sources of error may lead to missed opportunities to identify larger, unintentional changes in how equipment and automated tools are being used. SPC can produce individualized results per contour, as shown in Table 1. Across different TPS/algorithm combinations and anatomical sites, the mean and range of the control limits varied from tenths of a percent to greater

than 5%. This variability arose from several factors, including, but not limited to, TPS/algorithm dose calculation differences and calibration of machine parameters.

A phantom study by Alghamdi and Tajaldeen³³ showed that CCC had better dose agreement in out-of-field points compared to AAA and Acuros, whereas it did not agree as well for in-field points of varying clinical densities. Similarly, our data showed, across all examined anatomical sites, that the largest differences between RayStation-CCC and Eclipse-Acurus were in the doses to primary targets, followed by OARs located proximally to the primary target (e.g., oral cavity contour for head and neck, left and right femoral heads for cervix). The doses then began to equalize toward the distal OARs (e.g., brain contour for head and neck, heart contour for chest wall). Particularly for targets near the surface in head and neck cases, this variability in SPC control limits may be attributed in part to differences in dose calculations near the surface and to differing body contours between Eclipse and RayStation.¹² These factors also may explain why the dose differences in chest wall cases and head and neck cases were similar among the various contours, but were not as pronounced in the cervix cases, in which the dose differences were more uniform across contours. In addition, this might also provide an explanation as to why we see the opposite effect of overall dose in Raystation-CCC versus Eclipse-Acurus for cervix plans compared to head and neck and chest wall plans.

The dependency of the plan on the choice of machine parameters can also play a role in determining acceptable control limits. We found that as DLG or MLC-TF increased, both the mean and range of control limits changed for some contours. Specifically, we observed a smaller range between UCL and LCL with AAA than with Acuros. This is unsurprising because the RPA plan also used AAA, though calibrated to different machine parameters. However, this finding also demonstrates the need to set variable limits when monitoring a system like the RPA; different RT facilities use different TPS/algorithm combinations. Specifically, Martin-Martin et al.³⁴ found that AAA and Acuros can differ by up to 6.3% if not properly characterized for VMAT head and neck flattening filter-free RT cases—well beyond a clinical threshold, though this does not necessarily reflect the quality of the plan.

As the RPA is multi-institutional, it is imperative to design a flexible QA system that can adapt to differing setups and machine configurations. Toward this end, the mean and control limits depend on setup of the user's TPS; therefore, the SPC limits are institution specific. It is important to emphasize this fact of the SPC limits being institution specific, because the use of SPC is for the purpose of analyzing the process that invariably differs, sometimes largely, between institutions. Specifically, agreement of their calculations with the RPA will depend on how their TPS is commissioned, so SPC cannot be

expected to identify commissioning errors, but can be expected to identify changes in the TPS.

Furthermore, we observed that not all machine parameters have the same impact across all contours. In particular, we observed that primary targets were the most affected when DLG was increased, in terms of both means and control limits. In contrast, increases in MLC-TF tended to affect all contours similarly. In addition, AAA plans were less affected by changes in these machine parameters than were Acuros plans, for all contours. Nonetheless, Acuros and AAA became more similar as both DLG and MLC-TF increased, specifically, when DLG was 0.2 cm and MLC-TF 0.02, the values used by the RPA machine. However, we did observe outliers (e.g., eye, lens, and optic nerves) that skewed the value of MLC-TF in which Acuros and RPA were equal. The increased sensitivity of dose changes to relatively small contours could explain why these contours were outliers and justify their removal from the calculation.

In this study, we investigated monitoring of mean dose differences. This choice was initially made so that anticipated differences between treatment planning systems (especially near the skin surface) would not overly affect our ability to monitor changes, particularly gross errors. Also, we did show that SPC can identify changes in process (e.g., in the user's planning system); however, one limitation of this work is that we did not investigate other metrics, such as several dose volume histogram metrics, maximum dose, or dose to the hottest 1 cm³. These other metrics might be more adept, when monitored through SPC, at identifying errors within the process and increasing sensitivity and specificity of our process QA. Further investigation is needed to identify the most appropriate parameter.

5 | CONCLUSIONS

This study used SPC to determine the differences in mean dose and control limits between RPA and two separately commissioned TPSs. For head and neck and chest wall cases, most contours had only small mean differences under different TPS/algorithm combinations and machine parameters. Because it can account for the effects of varying control limits and means for different users and different planning approaches, SPC provides a flexible and useful QA tool for monitoring a complex autocontouring and autoplanning system such as the RPA.

AUTHOR CONTRIBUTIONS

All the authors have contributed to the design, acquisition, analysis, or interpretation of the data for this work. In addition, all authors helped in the writing or revision of the manuscript, and approved the final draft for submission.

ACKNOWLEDGMENTS

This work was carried out in collaboration with the Radiation Planning Assistant Group, which has funding from the NCI, CPRIT, Wellcome Trust, and Varian Medical Systems.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

- Netherton TJ, Cardenas CE, Rhee DJ, Court LE, Beadle BM. The emergence of artificial intelligence within radiation oncology treatment planning. *Oncology*. 2020;99(2):124-134. <https://doi.org/10.1159/000512172>
- Kisling K, Johnson JL, Simonds H, et al. A risk assessment of automated treatment planning and recommendations for clinical deployment. *Med Phys*. 2019;46(6):2567-2574. <https://doi.org/10.1002/mp.13552>
- Court LE, Kisling K, McCarroll R, et al. Radiation planning assistant - a streamlined, fully automated radiotherapy treatment planning system. *J Vis Exp*. 2018;134. <https://doi.org/10.3791/57411>
- McCarroll RE, Beadle BM, Balter PA, et al. Retrospective validation and clinical implementation of automated contouring of organs at risk in the head and neck: a step toward automated radiation treatment planning for low- and middle-income countries. *J Glob Oncol*. 2018;4:1-11. <https://doi.org/10.1200/jgo.18.00055>
- Rhee DJ, Cardenas CE, Elhalawani H, et al. Automatic detection of contouring errors using convolutional neural networks. *Med Phys*. 2019;46(11):5086-5097. <https://doi.org/10.1002/mp.13814>
- Olanrewaju A, Court LE, Zhang L, et al. Clinical acceptability of automated radiation treatment planning for head and neck cancer using the Radiation Planning assistant. *Pract Radiat Oncol*. 2021;11(3):177-184. <https://doi.org/10.1016/j.prro.2020.12.003>
- Kisling K, Zhang L, Shaitelman SF, et al. Automated treatment planning of postmastectomy radiotherapy. *Med Phys*. 2019;46(9):3767-3775. <https://doi.org/10.1002/mp.13586>
- Kisling K, Zhang L, Simonds H, et al. Fully automatic treatment planning for external-beam radiation therapy of locally advanced cervical cancer: a tool for low-resource clinics. *J Glob Oncol*. 2019;5:1-9. <https://doi.org/10.1200/jgo.18.00107>
- Rhee DJ, Jhingran A, Kisling K, Cardenas C, Simonds H, Court L. Automated radiation treatment planning for cervical cancer. *Semin Radiat Oncol*. 2020;30(4):340-347. <https://doi.org/10.1016/j.semradonc.2020.05.006>
- Rhee DJ, Jhingran A, Rigaud B, et al. Automatic contouring system for cervical cancer using convolutional neural networks. *Med Phys*. 2020;47(11):5648-5658. <https://doi.org/10.1002/mp.14467>
- Han EY, Cardenas CE, Nguyen C, et al. Clinical implementation of automated treatment planning for whole-brain radiotherapy. *J Appl Clin Med Phys*. 2021;22(9):94-102. <https://doi.org/10.1002/acm2.13350>
- Lee J, Park JM, Wu H-G, Kim JH, Ye S-J. The effect of body contouring on the dose distribution delivered with volumetric-modulated arc therapy technique. *J Appl Clin Med Phys*. 2015;16(6):365-375. <https://doi.org/10.1120/jacmp.v16i6.5810>
- Breen SL, Moseley DJ, Zhang B, Sharpe MB. Statistical process control for IMRT dosimetric verification. *Med Phys*. 2008;35(10):4417-4425. <https://doi.org/10.1118/1.2975144>
- Pawlicki T, Yoo S, Court LE, et al. Process control analysis of IMRT QA: implications for clinical trials. *Phys Med Biol*. 2008;53(18):5193-5205. <https://doi.org/10.1088/0031-9155/53/18/023>
- Gagneur JD, Ezzell GA. An improvement in IMRT QA results and beam matching in LINACS using statistical process control. *J Appl Clin Med Phys*. 2014;15(5):190-195. <https://doi.org/10.1120/jacmp.v15i5.4927>
- Gérard K, Grandhaye J-P, Marchesi V, Kafrouni H, Husson F, Aletti P. A comprehensive analysis of the IMRT dose delivery process using statistical process control (SPC). *Med Phys*. 2009;36(4):1275-1285. <https://doi.org/10.1118/1.3089793>
- Pawlicki T, Yoo S, Court LE, et al. Moving from IMRT QA measurements toward independent computer calculations using control charts. *Radiother Oncol*. 2008;89(3):330-337. <https://doi.org/10.1016/j.radonc.2008.07.002>
- Xiao Q, Bai S, Li G, et al. Statistical process control and process capability analysis for non-normal volumetric modulated arc therapy patient-specific quality assurance processes. *Med Phys*. 2020;47(10):4694-4702. <https://doi.org/10.1002/mp.14399>
- Sanghangthum T, Suriyapee S, Srisatit S, Pawlicki T. Statistical process control analysis for patient-specific IMRT and VMAT QA. *J Radiat Res*. 2012;54(3):546-552. <https://doi.org/10.1093/jrr/rrs112>
- Palaniswaamy G, Scott Brame R, Yaddanapudi S, Rangaraj D, Mucic S. A statistical approach to IMRT patient-specific QA. *Med Phys*. 2012;39(12):7560-7570. <https://doi.org/10.1118/1.4768161>
- Pardo J, González J, Aymar N, et al. Statistical control of processes applied to geometric uncertainties for CTV expansion margins determination in prostate cancer patients treated with VMAT: a prospective study in 57 patients. *Clin Transl Oncol*. 2020;23(6):1078-1084. <https://doi.org/10.1007/s12094-020-02493-6>
- Moore SJ, Herst PM, Louwe RJW. Review of the patient positioning reproducibility in head-and-neck radiotherapy using statistical process control. *Radiother Oncol*. 2018;127(2):183-189. <https://doi.org/10.1016/j.radonc.2018.01.006>
- Létourneau D, Wang A, Amin MN, et al. Multileaf collimator performance monitoring and improvement using semiautomated quality control testing and statistical process control. *Med Phys*. 2014;41(12):1217-13. <https://doi.org/10.1118/1.4901520>
- Meyers SM, Balderson MJ, Létourneau D. Evaluation of Elekta agility multi-leaf collimator performance using statistical process control tools. *J Appl Clin Med Phys*. 2019;20(7):100-108. <https://doi.org/10.1002/acm2.12660>
- Binny D, Aland T, Archibald-Heeren BR, Trapp JV, Kairn T, Crowe SB. A multi-institutional evaluation of machine performance check system on treatment beam output and symmetry using statistical process control. *J Appl Clin Med Phys*. 2019;20(3):71-80. <https://doi.org/10.1002/acm2.12547>
- Shiraishi S, Grams MP, Fong de los Santos LE. Image-guided radiotherapy quality control: Statistical process control using image similarity metrics. *Med Phys*. 2018;45(5):1811-1821. <https://doi.org/10.1002/mp.12859>
- Wang H, Xue J, Chen T, et al. Adaptive radiotherapy based on statistical process control for oropharyngeal cancer. *J Appl Clin Med Phys*. 2020;21(9):171-177. <https://doi.org/10.1002/acm2.12993>
- Fraass B, Doppke K, Hunt M, Kutcher G, Starkschall G, Stern R, et al. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. *Med Phys*. 1998;25(10):1773-1829. <https://doi.org/10.1118/1.598373>
- Ma CM, Chetty IJ, Deng J, Faddegon B, Jiang SB, Li J, et al. Beam modeling and beam model commissioning for Monte Carlo dose calculation-based radiation therapy treatment planning: report of AAPM Task Group 157. *Med Phys*. 2020;47(1):e1-e18. <https://doi.org/10.1002/mp.13898>
- Wheeler DJ, Chambers DS. *Understanding Statistical Process Control*. SPC Press; 2010.
- Glenn MC, Peterson CB, Howell RM, Followill DS, Pollard-Larkin JM, Kry SF. Sensitivity of IROC phantom performance to radiotherapy treatment planning system beam modeling parameters

- based on community-driven data. *Med. Phys.* 2020;47:5250-5259. <https://doi.org/10.1002/mp.14396>
32. Glenn MC, Peterson CB, Howell RM, Followill DS, Pollard-Larkin JM, Kry SF. Sensitivity of IROC phantom performance to radiotherapy treatment planning system beam modeling parameters based on community-driven data. *Med Phys.* 2020;47(10):5250-5259. <https://doi.org/10.1002/mp.14396>
33. Alghamdi S, Tajaldeen A. Evaluation of dose calculation algorithms using different density materials for in-field and out-of-field conditions. *Exp Oncol.* 2019;41(1):46-52. <https://doi.org/10.32471/exp-oncology.2312-8852.vol-41-no-1.12529>
34. Martin-Martin G, Walter S, Guibelalde E. Dose accuracy improvement on head and neck VMAT treatments by using the ACUROS algorithm and accurate FFF beam calibration. *Rep Pract Oncol*

Radiother. 2021;26(1):73-85. <https://doi.org/10.5603/rpor.a2021.0014>

How to cite this article: Mehrens H, Douglas R, Gronberg M, Nealon K, Zhang J, Court L. Statistical process control to monitor use of a web-based autoplanning tool. *J Appl Clin Med Phys.* 2022;23:e13803. <https://doi.org/10.1002/acm2.13803>