Check for updates

Received: 21 June 2023 | Revised: 13 November 2023 | Accepted: 7 December 2023

DOI: 10.1002/mp.16907

RESEARCH ARTICLE

MEDICAL PHYSICS

Multi-institutional investigation into the robustness of intra-cranial multi-target stereotactic radiosurgery plans to delivery errors

Lauren May[1] | **Nicholas Hardcastle**[1,2] | **Victor Hernandez**[3] | **Jordi Saez**[4] | **Anatoly Rosenfeld**[1] | **Joel Poder**[1,5,6]

[1]Centre for Medical and Radiation Physics, University of Wollongong, North Wollongong, NSW, Australia

[2]Physical Sciences, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

[3]Department of Medical Physics, Hospital Universitari Sant Joan de Reus, IISPV, Tarragona, Spain

[4]Department of Radiation Oncology, Hospital Clínic de Barcelona, Barcelona, Spain

[5]St George Cancer Care Centre, St George Hospital, Kogarah, NSW, Australia

[6]School of Physics, University of Sydney, Camperdown, NSW, Australia

Correspondence
Lauren May, Faculty of Engineering & Information Sciences, University of Wollongong 4 G14, Northfields Ave, North Wollongong, NSW 2500, Australia.
Email: lmmp504@uowmail.edu.au

Abstract

Background: The use of modulated techniques for intra-cranial stereotactic radiosurgery (SRS) results in highly modulated fields with small apertures, which may be susceptible to uncertainties in the delivery device.

Purpose: This study aimed to quantify the impact of simulated delivery errors on treatment plan dosimetry and how this is affected by treatment planning system (TPS), plan geometry, delivery technique, and plan complexity. A beam modelling error was also included as context to the dose uncertainties due to treatment delivery errors.

Methods: Delivery errors were assessed for multiple-target brain SRS plans obtained through the Trans-Tasman Radiation Oncology Group (TROG) international treatment planning challenge (2018). The challenge dataset consisted of five intra-cranial targets, each with a prescription of 20 Gy. Of the final dataset of 54 plans, 51 were created using the volumetric modulated arc therapy (VMAT) technique and three used intensity modulated arc therapy (IMRT). Thirty-five plans were from the Varian Eclipse TPS, 17 from Elekta Monaco TPS, and one plan each from RayStation and Philips Pinnacle TPS. The errors introduced included: monitor unit calibration errors, multi-leaf collimator (MLC) bank offset, single MLC leaf offset, couch rotations, and collimator rotations. Dosimetric leaf gap (DLG) error was also included as a beam modelling error. Dose to targets was assessed via dose covering 98% of planning target volume (PTV) (D98%), dose covering 2% of PTV (D2%), and dose covering 99% of gross tumor volume (GTV) (D99%). Dose to organs at risk (OARs) was assessed using the volume of normal brain receiving 12 Gy (V12Gy), mean dose to normal brain, and maximum dose covering 0.03cc brainstem (D0.03cc). Plan complexity was also assessed via edge metric, modulation complexity score (MCS), mean MLC gap, mean MLC speed, and plan modulation (PM).

Results: PTV D98% showed high robustness on average to most errors with the exception of a bank shift of 1.0 mm and large rotational errors $\geq 1.0°$ for either the couch or collimator. However, in some cases, errors close to or within generally accepted machine tolerances resulted in clinically relevant impacts. The greatest impact upon normal brain V12Gy, mean dose to normal brain, and D0.03cc brainstem was found for DLG error in alignment with other recent studies. All delivery errors had on average a minimal impact across these

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

parameters. Comparing plans from the Monaco TPS and the Eclipse TPS, showed a lesser increase to V12Gy, mean dose to normal brain, and D0.03cc brainstem for Monaco plans ($p < 0.01$) when DLG error was simulated. Monaco plans also correlated to lower plan complexity. Using Spearman's correlation coefficient (r) a strong negative correlation ($r \leq -0.8$) was found between the mean MLC gap and dose to OARs for DLG errors.

Conclusions: Reducing MLC complexity and using larger mean MLC gaps is recommended to improve plan robustness and reduce sensitivity to delivery and modelling errors. For cases in which the calculated dose distribution or dose indices are close to the clinically acceptable limits, this is especially important.

KEYWORDS
intra-cranial stereotactic radiosurgery, linear accelerator-based radiotherapy, plan robustness, plan complexity, delivery errors

1 INTRODUCTION

Intra-cranial stereotactic radiosurgery (SRS) is being increasingly delivered with standard linear accelerators (linacs) using highly modulated techniques.[1–4] The use of standard linacs and multi-leaf collimators (MLCs) necessitates small fields, and in particular with multiple metastases often intensity modulation is utilized in order to achieve a sharp dose fall off and high conformity. Small fields and high modulation can be more susceptible to variations in delivery and beam modelling errors.[5,6]

A growth in the use of single-isocenter multiple-target (SIMT) SRS,[7–10] where the isocenter is usually located at the center of the target distribution, means that some targets may be at off-axis positions and the impact of rotational errors are accentuated and play a greater role in plan deliverability.[11,12] Also of interest with the recent dominance of using MLCs over cones as indicated by survey data from Australia and New Zealand,[4] are MLC related errors. Multiple cases of MLC errors occurring clinically have been reported in recent years.[13,14]

In addition to delivery errors, beam modelling errors may also be a significant consideration. An investigation into Imaging and Radiation Oncology Core (IROC) Houston audit results and corresponding survey data by Glenn et al.[15] found that atypical treatment planning system (TPS) beam modelling parameters were associated with lower phantom audit results.

This study aimed to analyze treatment plans from a multi-institutional planning challenge to quantify the impact of delivery errors on multi-target intracranial SRS plan dosimetry. To provide context on dose uncertainties throughout treatment, an example of the beam modelling error highlighted as most impactful by Glen et al.[15] was also simulated. We further assessed the impact of factors including TPS used, plan geometry, delivery techniques, and plan complexity on plan robustness to simulated errors.

2 METHODS

The impact of a variety of delivery errors were assessed for multiple-target brain SRS plans obtained through the Trans-Tasman Radiation Oncology Group (TROG) treatment planning challenge in 2018.[16] In this challenge, CT and RTSTRUCT files for a single patient were provided to all participants, who performed a treatment plan according to a pre-defined protocol and scoring matrix (see supplementary material). The challenge patient had five targets, each with a prescription of 20 Gy. The planning challenge provided planning target volumes (PTVs) only, with target dosimetry requirements based on this volume. GTV volumes were retrospectively created by isotropically contracting the PTV by 1 mm. The target properties are outlined in Table 1. Note that PTV5 was positioned with close proximity to the brainstem (Figure 1).

The overall workflow for this study is outlined in Figure 2.

2.1 Importing and recalculating plans

The RayStation TPS (v.10.1, RaySearch Laboratories, Stockholm, Sweden) was used for all calculations and

TABLE 1 Target distribution and properties.

Target	PTV volume (cc)	GTV volume (cc)	Minimum length along x-, y-, and z-axes (cm)	Distance from center of target distribution (cm)
PTV1	0.5	0.2	1.0	5.5
PTV2	0.4	0.1	0.9	6.7
PTV3	<0.1	<0.1	0.4	5.0
PTV4	2.8	1.8	1.6	2.0
PTV5	0.1	<0.1	0.6	5.5

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

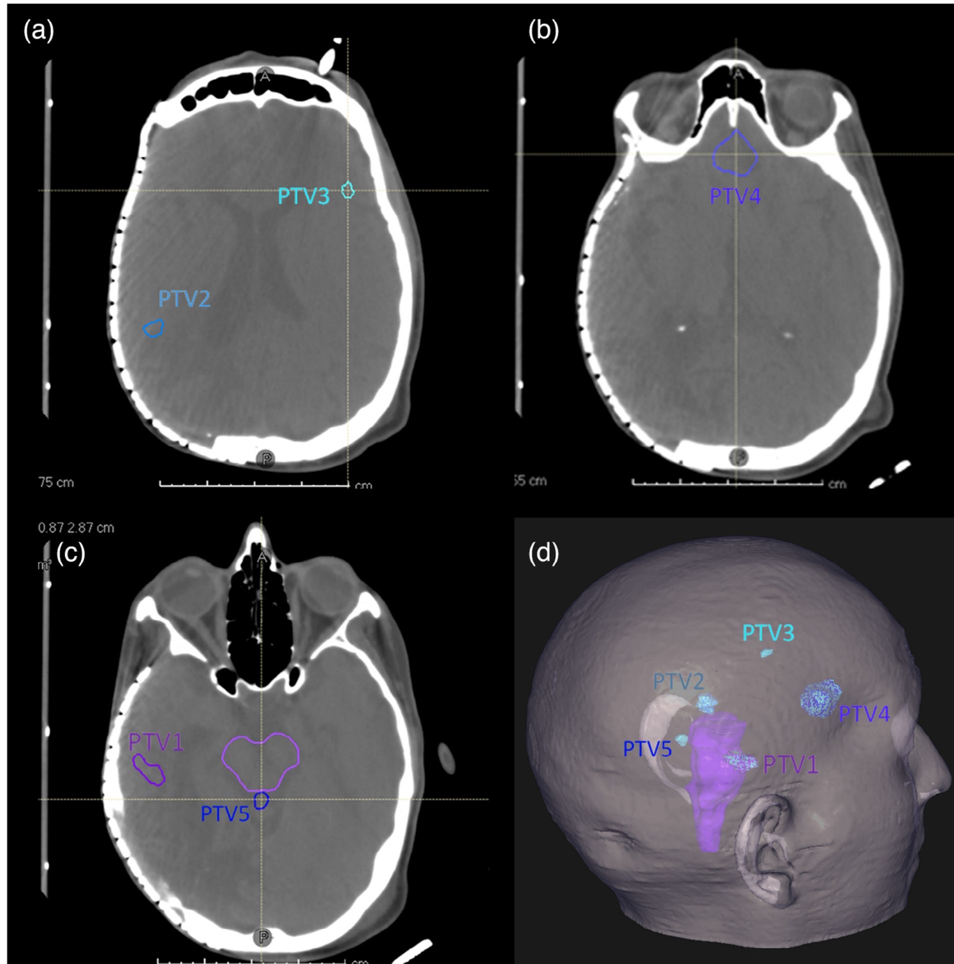

FIGURE 1 PTV locations relative to patient anatomy from patient CT scan for (a) PTV2 and 3, (b) PTV4 and (c) PTV1 and 5, with (d) showing a 3D render of all targets shown in blue, with the brainstem in purple.

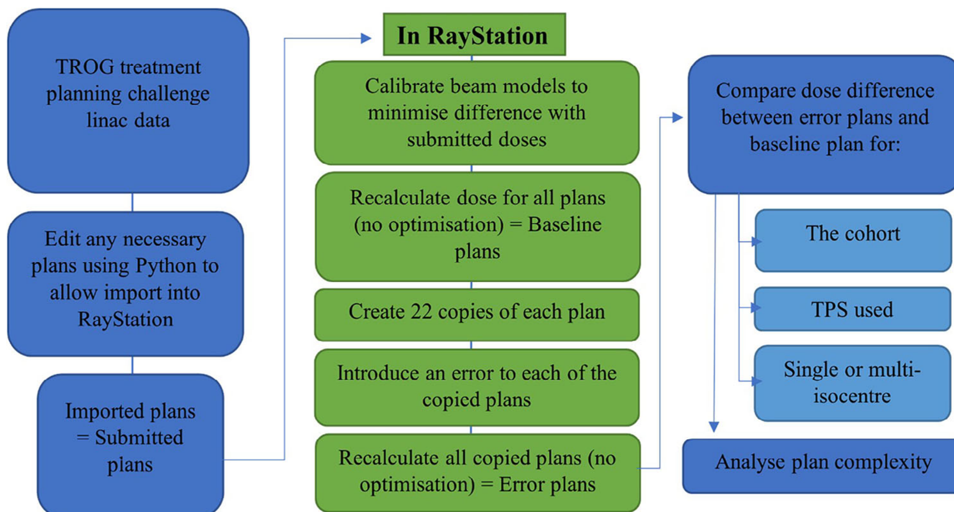

FIGURE 2 Workflow diagram for the assessment of introduced delivery errors across treatment plans.

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

assessment. A total of 88 treatment plans for conventional linear accelerators with MLCs were submitted for the challenge; a total of 54 were finally used in the study. Nine were excluded due to import errors either unable to be rectified or which required excessive modification, 16 due to not having access to relevant beam models and nine as outliers when comparing dose calculated within RayStation to submitted dose files.

The CT, RTSTRUCT, and RTDOSE DICOM data were imported into the TPS. Submitted DICOM RTPLAN files created by Pinnacle (Philips, Madison, WI, USA), Eclipse (Varian Medical Systems, Palo Alto, CA, USA) and RayStation TPSs were able to be imported without modification of the DICOM RT plan file. However, Monaco (Elekta CMS, Maryland Heights, MO, USA) submitted plans required editing in Python programming language (v.3.5.8, Python Software foundation, Fredericksburg, Virginia, USA). Most Monaco submitted plans contained VMAT arcs in which the gantry changed direction, which is not allowable in RayStation; these arcs were split into two at the direction change. The majority of Monaco submitted plans exceeded a 4° gantry spacing between control points for avoidance sectors of the arcs which is the maximum allowable spacing in RayStation. Therefore, additional control points were added as necessary with zero MUs distributed via linear interpolation between each existing control point in the modified DICOM RTPLAN files. Nine plans with maximum control point spacing of >20° were excluded due to the excessive number of control points which would be required.

Generic beam models for the Varian TrueBeam with the Millennium 120 leaf MLC (6MV flattening filter free (FFF) energy) and HD120 MLC (6FFF and 10FFF), and the Elekta Versa HD™ (6FFF) with the 160 Agility MLC were available in RayStation. Of the remaining 79 submitted plans, 16 were calculated by participating departments on beam models for machines/energies we did not have access to, which were the Varian Edge ($n = 2$); Varian Novalis Tx™ ($n = 3$); Varian Clinac iX ($n = 2$); Varian Trilogy ($n = 1$); Varian TrueBeam Millennium MLC 10FFF energy ($n = 2$); Elekta Synergy ($n = 2$); and Elekta Versa HD™ 8FFF ($n = 1$), 10FFF ($n = 2$), and 12FFF ($n = 1$). The remaining 63 submitted plans were recalculated on their respective beam models in RayStation, as described below.

Between TPSs there is variation in beam modelling, dose calculation algorithms, and sampling of volumes and dose. Therefore, there were differences between the dose calculated and submitted for the challenge by individual departments of participants and dose calculated in this study using generic beam models available in RayStation. In order to minimize this difference, beam modelling parameters source width, MLC offset, MLC gain, MLC curvature, leaf tip width, tongue and groove, and MLC transmission were initially set to match values contained in the supplementary data of Glenn et al.,[17] with the exception of parameters for the 6FFF TrueBeam with high definition (HD) MLC, as they were not included in this data. These values were instead initially set to match those outlined in Lee and Kim.[18] Each beam model was then manually adjusted to reduce the average difference across D2%, D98%, and V100% between submitted dose files and recalculated RayStation baseline dose to <10% (see Table S1). All models required the MLC offset to be altered with the largest adjustment a 0.070 cm reduction for the Versa HD model. The TrueBeam Millennium MLC beam model was the only one with altered source width reduced by 0.018 cm in both x- and y-parameter and the MLC transmission was reduced by 0.11% for the TrueBeam HDMLC 6FFF model. Nine outlying plans with an average difference across D2%, D98%, and V100% of >15% were excluded (Figure S1) resulting in a final dataset of 54 plans.

Submitted plans re-calculated in RayStation using adjusted beam models will henceforth be referred to as "baseline plans".

2.2 | Introducing errors

Copies of each baseline plan were created, which were then edited via scripting within RayStation using IronPython (v.3.4) to contain either a delivery or beam modelling error before being recalculated. The types of delivery errors included those that were considered in the recent study by Lehmann et al.[19] that proposed a novel auditing concept.

Errors included monitor unit calibration errors simulated via a change in field MUs of magnitudes −2.0%, −1.0%, +1.0%, and +2.0%. MLC related errors included an MLC bank offset (both banks shifted in the same direction) of 0.5 and 1.0 mm with the magnitude of shift in alignment with Rangel and Dunscombe[20] who introduced a 1.0 mm shift in a single bank. A single central MLC leaf located near central axis was offset by 1.0 and 2.0 mm with this magnitude based on a reported visually observed 1.3 mm error in a single MLC leaf pair position.[14] A central leaf was selected as this area is less likely to be shielded by secondary collimator jaws throughout treatment. Leaf 30 was selected for TrueBeam beam models with 60 leaf pairs and leaf 40 for the Versa HD beam model with 80 leaf pairs.

Rotational errors were also included with couch angle shifts of −1.0°, −0.5°, +0.5°, and +1.0°, collimator angle shifts of −2.0°, −1.0°, −0.5°, +0.5°, +1.0°, and +2.0° as well as a combination of the two with magnitudes of −0.5° and +0.5° with these shifts informed by Pudsey et al.[12]

Eclipse beam modelling contains a "dosimetric leaf gap" parameter (DLG) that accounts for the additional transmission through the leaf tip and for the specific leaf positioning calibration. Monaco, Pinnacle, and

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

RayStation instead use multiple MLC parameters to model what is hereafter referred to as DLG. Rather than altering any of these parameters, DLG errors were simulated by opening both MLC banks by 0.2 and 0.5 mm which is equivalent to increasing the DLG by twice this distance (0.4 and 1.0 mm, respectively) regardless of the TPS used. The DLG was only simulated to increase not decrease as moving the MLC banks towards each other might result in an error due to an infeasible or overlapping MLC pattern.

2.3 | Data analyses

The percentage difference between each baseline plan and error plan (calculated as [error—baseline]/baseline) was assessed using near-minimum (D98%) and near-maximum (D2%) for each PTV as recommended in the international commission on radiation units and measurements (ICRU) report 91[21] and D99% for GTVs. Dose to organs at risk (OARs) was also assessed using volume of normal brain receiving 12 Gy (V12Gy), mean dose to normal brain and maximum dose covering 0.03cc (D0.03cc) of brainstem. Plans were separated into single and multi-isocenter as well as grouped by TPS for comparison.

Complexity of the MLC pattern was analyzed using a range of metrics: edge metric, modulation complexity score (MCS), mean MLC gap, mean MLC speed and plan modulation (PM). These parameters have previously been defined and were calculated using PlanAnalyzer implemented in Matlab.[22]

Significance when comparing two subsets of data was tested using the Mann–Whitney U test, otherwise correlation between variables was tested using Spearman's correlation coefficient. It should be noted, as multiple hypotheses were tested and the p-values were not corrected (e.g., using Bonferroni correction) the likelihood of incorrectly rejecting a null hypothesis may be increased and therefore the p-values obtained through these statistical tests act as identifiers of standout comparisons rather than global significant differences.

3 | RESULTS

Of the final dataset of 54 plans, 51 were created using the volumetric modulated arc therapy (VMAT) technique and three used intensity modulated radiation therapy (IMRT). Thirty-five plans were from the Varian Eclipse TPS, 17 from Elekta Monaco TPS, and one plan each from RayStation and Philips Pinnacle. Plan properties are shown in Table 2.

Initial absolute values for OAR dosimetry for the baseline plans as calculated in RayStation are shown in Table 3.

TABLE 2 Properties of obtained plans.

Treatment planning system	Machine	MLC	Energy	VMAT plans	IMRT plans	Total plans
Eclipse	TrueBeam	Millennium	6FFF	5	0	5
		HD120	6FFF	24	2	26
			10FFF	4	0	4
Monaco	Versa HD	Standard	6FFF	17	0	17
Pinnacle	TrueBeam	Millennium	6FFF	0	1	1
RayStation	Versa HD	Agility	6FFF	1	0	1

TABLE 3 Absolute values for OAR dosimetry for baseline plans as calculated in RayStation.

OAR parameter	Median[range]
Normal brain V12Gy	11.0 [6.7–58.7]cc
Brainstem D0.03cc	12.7 [7.2–19.7]Gy
Normal brain mean dose	2.3 [1.4–4.1]Gy

3.1 | Impact of introduced errors across all plans

3.1.1 | Target dose

Most errors resulted in clinically significant reductions in target coverage (PTV D98%) with some plans from the cohort having reductions of >10% for most errors (Figure 3); however, median reduction across the cohort remained relatively low. Excluding a bank shift of 1.0 mm magnitude and large rotational errors of ≥1.0° for which the greatest impacts were seen, all other errors had 119 PTVs (of 4050 included in this analysis) with PTV coverage reduced by >5%, while the median[range] change in coverage was 0.0%[−19.5–32.7]%. For a 1 mm bank shift, change in PTV D98% was −5.0%[−69.3–6.5]% and for large rotational errors median[range] change in PTV D98% was −3.3%[−46.1–5.2]. GTV coverage (GTV D99%) and PTV D2% followed a similar trend (Figure S2).

3.1.2 | Dose to organs at risk

With the exclusion of DLG errors, 35 plans had an increase of >10% in normal brain V12Gy and the median[range] change was 0.08%[−34.2–20.6] (Figure 3). The change in brainstem near max was 0.1%[−19.1–54.9]% with 40 plans of 1080 showing an increase of >10%. Though median values were relatively low there were nine plans with an increase of >20%, seven of which were from 2.0 mm leaf errors, one from a 1.0 mm bank shift and in one case a 1.0 mm leaf error showed a 23.6% increase to brainstem D0.03cc. More consistently greater impacts were seen for DLG

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

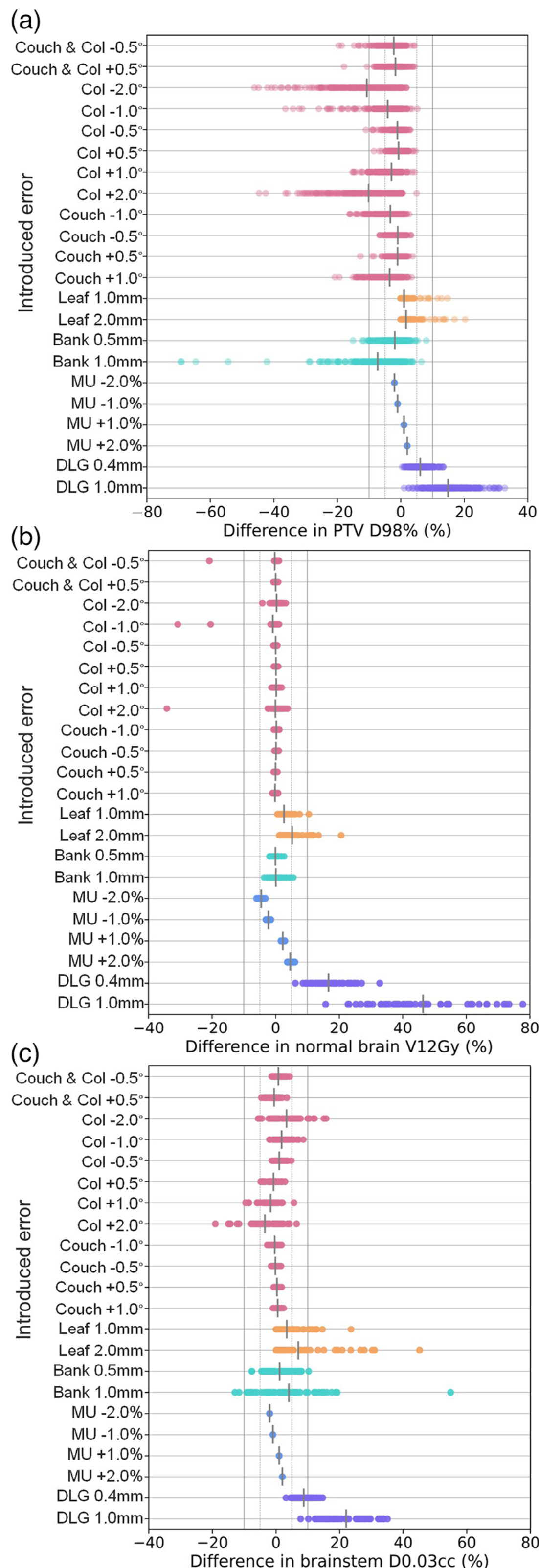

FIGURE 3 Relative change between baseline plans with dose calculated in RayStation and error plans calculated as [error—baseline]/baseline for (a) PTV D98%, (b) normal brain V12Gy, and (c) brainstem D0.03cc. Each data point is represented as a dot and the vertical lines represent the mean value.

error as a 0.4 mm DLG error resulted in an increase in V12Gy of 15.7%[6.2–32.8]% and brainstem near max of 8.3%[3.1–14.7]%. The impact upon normal brain mean dose showed similar trends though with minimal outlying values (Figure S3).

Using the median quantitative values for OAR dosimetry across the cohort of baseline plans (Table 3), the median 15.7% increase in normal brain V12Gy with a 0.4 mm DLG error corresponds to a 1.7cc increase. For brainstem near max, this would be a 1.1 Gy increase and for normal brain mean dose a 0.2 Gy increase.

3.2 | Single- and multi-isocenter plans

Of the obtained dataset, 42 plans were created using a single isocenter; of the remaining 12 plans, 11 had between 2 and 5 isocenters and one plan had eight isocenters.

3.2.1 | Target dose

Shown in Figure 4(a) is a comparison of change in PTV D98% for single- and multi-isocenter plans for a selection of errors. Rotational errors resulted in a smaller reduction in PTV D98% and GTV D99% ($p < 0.01$) for multi-isocenter plans compared to single-isocenter plans for all rotations and magnitudes (Figure S4; Table S2). In contrast to rotational errors, shifting the MLC banks in the same direction resulted in a lesser median reduction of PTV D98% and GTV D99% for single-isocenter plans ($p < 0.01$); however, the range of values was greater.

3.2.2 | OAR dosimetry

There was little difference in OAR dosimetry between single- and multi-isocenter plans when errors were introduced. Although the impact on OARs were different ($p < 0.05$) for some rotational errors, most bank shift magnitudes, single leaf errors and a 0.4 mm DLG error for brainstem near max, the magnitude of difference was clinically insignificant (Figure 4; Figure S5 and Table S3).

3.3 | Comparison of treatment planning systems

The comparison between TPSs was limited to 6FFF single-isocenter VMAT plans from Eclipse ($n = 21$) or Monaco ($n = 17$). Due to only obtaining one plan from RayStation and Pinnacle, these were not included.

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

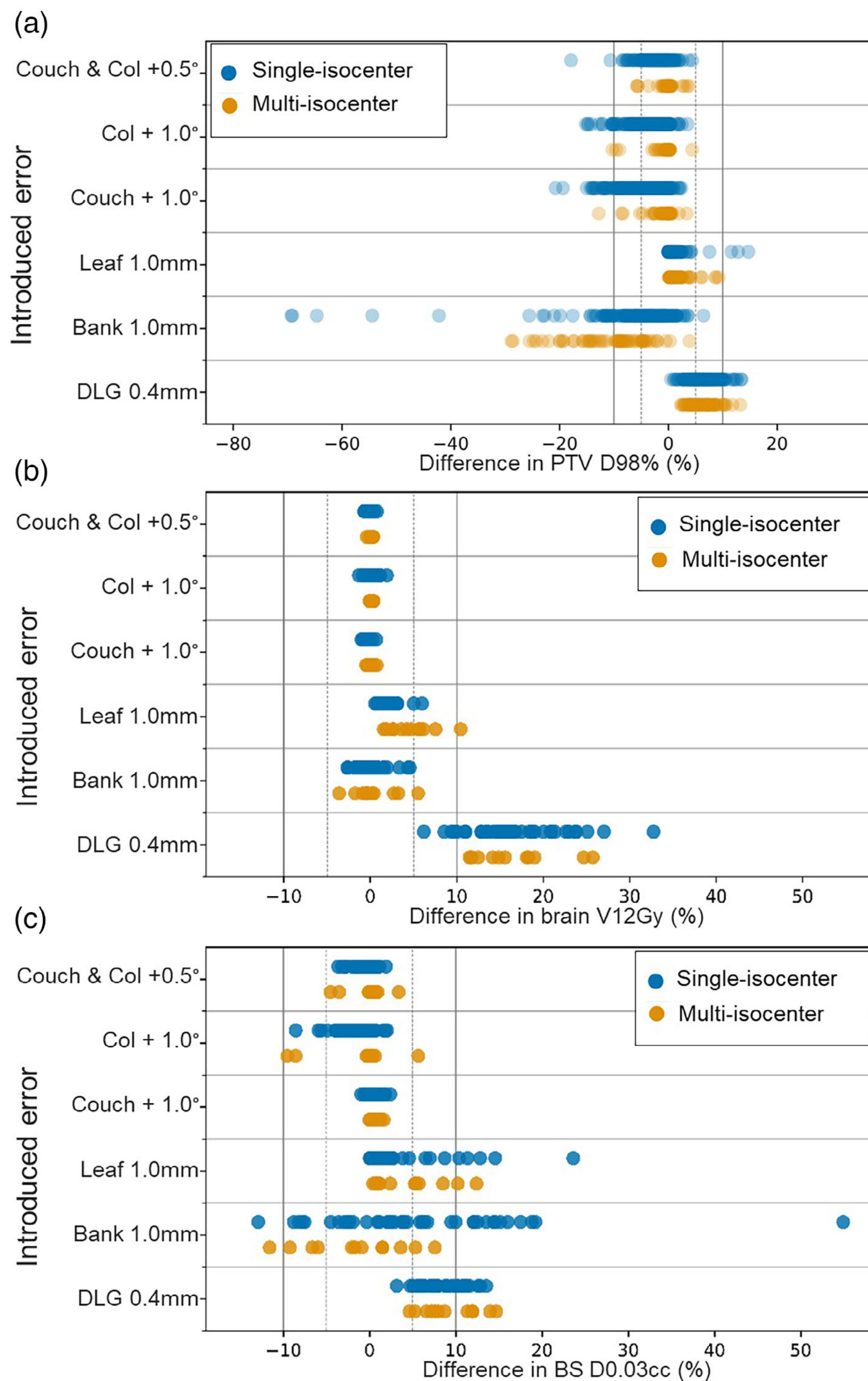

FIGURE 4 Comparison of the change from baseline plans when errors are introduced calculated as [error—baseline]/baseline, between single-isocenter (blue/upper) and multi-isocenter (gold/lower) plans for (a) PTV D98%, (b) normal brain V12Gy, and (c) brainstem (BS) D0.03cc. Each dot is data for one plan.

3.3.1 | Target dosimetry

Both couch and collimator rotational errors resulted in a greater reduction in GTV D99% for Eclipse plans compared to Monaco plans for all magnitudes of rotation ($p < 0.05$) (Figure S6; Table S4). The greatest difference in median reduction across these errors was 3.7% for a +2.0° collimator rotation. Change in PTV D98% was also lower ($p < 0.05$) for Eclipse plans with the exception of a +0.5° couch rotation ($p = 0.14$), −0.5° couch rotation ($p = 0.05$) and a combined +0.5° couch and collimator rotation ($p = 0.09$) (Figure 5a). The greatest difference across these errors was 5.3% for a −2.0° collimator shift. A 0.5 mm bank shift had no

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

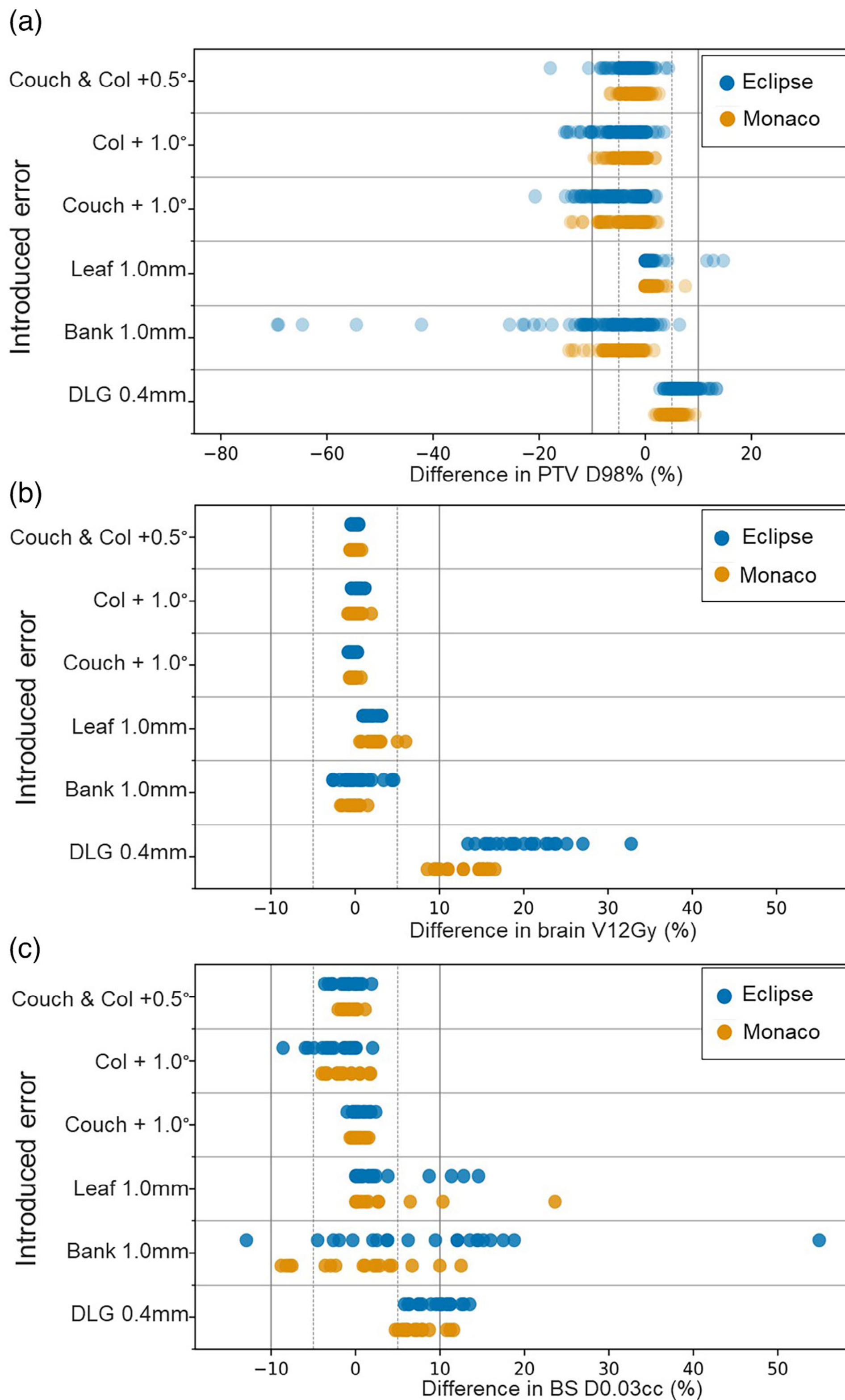

FIGURE 5 Comparison of the change from baseline plans when errors are introduced calculated as [error—baseline]/baseline, between plans created in Eclipse (blue/upper) and Monaco(gold/lower) for (a) PTV D98%, (b) normal brain V12Gy, and (c) brainstem (BS) D0.03cc. Each dot is data for one plan.

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 A comparison of complexity metrics for single-isocenter 6FFF VMAT plans made in the Monaco TPS and the Eclipse TPS.

Complexity metric	Mann–Whitney U	*p* value	Eclipse	Monaco
PM	1925	<0.01	0.98[0.95–0.99]	0.99[0.97–0.99]
Edge metric	375	<0.01	0.84[0.55–1.08]	0.52[0.43–0.63]
Mean MLC speed	325	<0.01	0.30[0.05–0.59] cm/s	0.05[0.01–0.14] cm/s
MCS	100	<0.01	0.03[0.02–0.08]	0.15[0.05–0.26]
Mean MLC gap	0	<0.01	3.70[2.99–6.17] mm	7.38[6.22–11.93] mm

significant difference in impact upon target coverage between Eclipse and Monaco plans; however, a 1.0 mm bank shift resulted in a greater reduction in GTV D99% for Eclipse plans over Monaco, though the magnitude of this difference was only 1.9%.

3.3.2 | OAR dosimetry

The simulated DLG error showed a greater increase in OAR dosimetry for Eclipse plans compared to Monaco plans ($p < 0.01$) (Figure 5; Figure S-7, Table S5). Normal brain V12Gy had a difference in medians for a 0.4 mm DLG error of 5.4% and brainstem D0.03cc of 3.0%. The difference in impact of MLC bank shifts between the two TPSs was statistically significant for brainstem D0.03cc ($p < 0.05$) as was the impact of large rotational errors of ≥1.0° magnitude where despite statistical significance the difference in clinical impact upon OARs between the two TPSs w negligible.

3.3.3 | Plan complexity

Monaco plans showed an overall lower complexity for all metrics compared to Eclipse plans except for PM which differed by only 0.01. Monaco plans had a lower edge metric and mean MLC speed as well as greater values for MCS and mean MLC gap with an inverse relationship to complexity (Table 4). For complexity metrics across all plans see Table S6, supplementary data.

Correlation was assessed using Spearman's correlation coefficient between complexity metrics and impact of errors on target dosimetry (Table S7).

The only strong correlation with target coverage ($0.6 < |r| < 0.8$) was an inverse correlation between mean MLC gap and PTV D98% and GTV D99% when DLG errors of both a 0.4 and 1.0 mm magnitude are introduced. However, it should be noted that the simulated DLG error opened the MLC leaves therefore increasing these parameters and not compromising PTV or GTV coverage. The number of isocenters had a moderate correlation ($0.4 < r < 0.6$) with the impact upon target coverage for bank shifts, some rotational errors and single leaf offsets. Moderate correlation was also found for MCS and mean MLC gap with target coverage for some rotational errors, and moderate inverse correlation for PM and target coverage for some rotational errors.

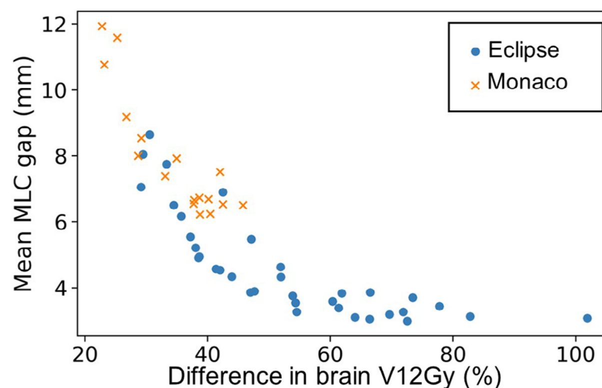

FIGURE 6 Scatter plot showing the relationship between mean MLC gap and the change in normal brain V12Gy when a DLG error of 1.0 mm is applied. Each point represents data for one plan with the gold crosses representing Monaco plans and the blue dots representing Eclipse plans.

Spearman's correlation coefficient showed correlation between some complexity metrics and impact of errors on dose to OARs particularly for DLG errors, with plans with higher complexity having, in general, larger impacts (Table S8). The strongest correlation occurred for DLG errors of both 0.4 and 1.0 mm magnitude with a very strong inverse correlation ($r < -0.8$, $p < 0.01$) between mean MLC gap and the variation in V12Gy and mean dose to normal brain. This trend can be seen in Figure 6. A strong inverse correlation was found ($-0.8 < r < -0.6$, $p < 0.01$) between mean MLC gap and brainstem near max when DLG errors are introduced as well as between MCS and normal brain V12Gy. There was a strong positive correlation ($0.6 < r < 0.8$, $p < 0.01$) between the edge metric and normal brain mean, V12Gy and brainstem D0.03cc. Finally, leaf errors showed a strong positive correlation between MCS and mean dose to normal brain for a 1.0 mm leaf shift.

4 | DISCUSSION

The robustness of plans to delivery errors in the form of rotational errors, MLC bank shifts, single MLC leaf errors, and dose calculation errors were evaluated for target coverage and dose to normal brain and the

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

brainstem across the entire dataset. The impact of DLG error was also included as a context of dose uncertainties with beam modelling errors. The effect of using single- or multi-isocenter plan geometry was investigated regarding the impact of these errors. The role the TPS plays with plan complexity and plan robustness was also assessed.

Target coverage was shown to be relatively robust to introduced errors when analyzing data across the cohort. However, even with the exclusion of a large bank shift (1 mm) and rotational errors $>0.5°$ magnitude, a small proportion of PTVs (119 of 4050) had a reduction in PTV D98% of $>5\%$. This indicates that while overall robustness may be high, clinically relevant dose errors can still be produced for delivery errors within or close to accepted machine tolerances.

As with target coverage, dose to normal brain and the brainstem showed high levels of robustness to most introduced errors across the cohort with the greatest impact seen for DLG errors. This is in alignment with recent findings reported by Glenn et al.[15] where data from the IROC Houston phantom audits and corresponding survey data were analyzed and it was found that atypical DLG values were more likely to have poor performing or failing audit results. It has also been shown in a previous study that dosimetric impact of positive and negative DLG variations are similar and quite symmetrical.[23]

It should be noted that, although single MLC leaf error showed on average a low impact on dose to normal brain and brainstem, some plans showed increases in D0.03cc brainstem of up to 23% for 1.0 mm error (Figure 3). No strong correlation was found between brainstem D0.03cc and plan geometry, TPS or complexity metrics when single leaf shifts were applied, limiting the ability to predict when a plan may be susceptible to this kind of error. Dependent on the configuration of beam apertures of individual treatment plans, such errors may also be impactful in terms of target coverage. Single leaf errors of this magnitude have been reported to have occurred clinically as in Neal et al.[14] where a single leaf offset of 1.3 mm was observed. Therefore, MLC QA must be maintained at a high standard to limit the occurrence of these errors.

One limitation of this work is the variation between beam models from participating departments and those used within this study, which required analyses to be conducted using relative change in dosimetric indices rather than quantitative values or pass/fail thresholds. The effect of using qualitative indices can be seen where a seemingly large median increase in normal brain V12Gy was found for a 0.4 mm DLG error of 15.7%. However, the median quantitative value across all baseline plans for V12Gy was 11.0 cc, therefore with the average increase applied the quantitative value this corresponds to a total volume of 12.7 cc, well within the recently suggested acceptable limit to avoid brain necrosis of 5–10 cc per target.[11] It should also be considered that the minor dosimetric changes to plans as a result of non-optimal beam modelling, that is, different slope of the penumbra or conformity, could alter the plan performance.

Conversely, a smaller median increase was observed for near max dose to brainstem indicated by D0.03cc brainstem with a 0.4 mm DLG error of 8.3%. However, the median quantitative value across the dataset was already 12.7 Gy, with recommended guidelines being $\leq$12.5 Gy.[11] Therefore, the median increase results in a value even further outside of acceptable limits. This can be attributed to the patient geometry in this study which included a target (PTV5, Figure 1c) in close proximity to the brainstem, and resulted in plans with relatively large values for brainstem D0.03cc compared to other OAR dose parameters including normal brain V12Gy. Results from this study indicate that the impact upon OARs for most delivery and beam modelling errors could be clinically relevant, especially in situations where patient geometry results in difficult planning situations in which certain parameters are already close to acceptable limits.

The use of single-isocenter plans has become common practice among departments with the majority of plans included in this study (78%) being single-isocenter, which it should be noted may have some effect on statistical power. As expected, comparison of single- and multi-isocenter plans saw a lesser impact of rotational errors upon target coverage for multi-isocenter plans. This is due to the increased off-axis distance of targets in single-isocenter plans accentuating the rotational shifts. However, all other errors showed no significant lower robustness to errors for single-isocenter plans. These findings support the continued use of single-isocenter plans for multi-target plans as long as appropriate QA for rotational errors is undertaken.

A comparison of the Monaco TPS and Eclipse TPS showed a greater robustness to errors for Monaco plans over Eclipse. There was less reduction in target coverage for Monaco plans compared to Eclipse with rotational errors (Figure 5); however, the magnitude of the difference between the two TPSs was relatively small. Similarly, a 1.0 mm bank shift had a greater reduction in GTV D99% for Eclipse plans; however, the magnitude of difference was only 1.9%.

DLG error showed a greater increase in mean dose to normal brain, normal brain V12Gy, and brainstem near max for Eclipse plans (Figure 5). As with target dosimetry, the magnitude of difference in medians between the two TPSs is relatively small with the median increase in brainstem D0.03cc for a 0.4 mm DLG error being 3.0% greater for Eclipse plans over Monaco plans. However, this magnitude of difference may be clinically relevant in situations where parameters are already approaching acceptable limits.

Although this limited the possibility of analyses with TPS used, the low number of both Pinnacle and RayStation submissions available for the dataset is

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

comparable to findings of other studies.[15,17] The vast majority of IROC Houston auditing data in Glenn et al.[15] were from Eclipse users (78.4%), followed by Pinnacle at 11.9%, and RayStation at only 9.8%, noting that Monaco was not included. In the current study, the differences between the robustness of Eclipse and Monaco plans to DLG errors are supported by the aforementioned study by Glenn et al.,[15] as atypical DLG values showed a higher correlation to lower audit scores for Eclipse plans, although this was in comparison to RayStation and Pinnacle. This difference may be attributed to the high level of MLC complexity favored by the optimizer within Eclipse. Monaco plans had a lower overall complexity as measured by a range of complexity parameters ($p < 0.01$). Recently, Hernandez et al.[24] highlighted plan complexity as an important consideration when assessing plan quality. Findings of the current study further support this as Spearman's correlation coefficient showed strong correlation between some complexity parameters and the robustness of plans to errors. The strongest correlation found was between the impact of the DLG error and mean MLC gap. This indicates that the higher robustness seen for Monaco plans to DLG errors is likely as a result of the TPS optimizer preference of larger MLC gaps and lower MLC complexity than Eclipse. This is illustrated in Figure 6 depicting a correlation between mean MLC gap and change in normal brain V12Gy when DLG errors are applied with data from Eclipse and Monaco plans tending towards opposing areas of the plot. Further, mean MLC gap may be a good indicator of robustness to DLG errors which had the greatest impact upon OAR dosimetry of all introduced errors.

These findings indicate that for challenging plans where dosimetric indices are already close to acceptable limits, a preference for lower plan complexity and in particular larger MLC gaps may limit the impact of some beam modelling or delivery errors which may otherwise impact a plans acceptability. In these kinds of plans, in order to improve dose volume histograms (DVHs) it may be necessary to push the limits of the TPS, which typically increases plan complexity. Minimizing the number of MUs without sacrificing DVH quality may be helpful to increase the mean MLC gap and improve the robustness of treatment plans.

The errors investigated in this study were all system-related. As future work, an investigation into the robustness of plans to patient-related errors could also be considered including both patient positioning on the treatment couch as well as brain parenchyma distortion in response to intracranial pressure changes.

5 | CONCLUSIONS

A cohort of multi-institutional plans from a TROG treatment planning challenge in 2018 were used to assess the robustness of multi-target intracranial plans to delivery errors for target coverage and dose to normal brain and the brainstem, with DLG error simulated to provide context with beam modelling errors. The greatest impact upon dose to normal brain and brainstem was found for DLG errors in alignment with other recent studies. However, the clinical impacts of these errors remain relatively small and would be most relevant in situations where a challenging plan is already approaching acceptable limits. The effect of using single- or multi-isocenter plan geometry was investigated finding that with the exception of rotational errors which was to be expected, single-isocenter plans showed equal robustness to errors compared to multi-isocenter plans. The role the treatment planning system plays with plan complexity and plan robustness was also assessed, finding that plans with lower complexity and in particular greater mean MLC gaps had a higher robustness to DLG and rotational errors. Reducing MLC complexity and using larger mean MLC gaps is recommended to improve the plan robustness and reduce its sensitivity to delivery and modelling errors. This is especially important for cases in which the calculated dose distribution/dose indices are close to the clinically acceptable limits.

ACKNOWLEDGMENTS

We thank the Trans-Tasman Radiation Oncology Group for making the data used in this study available.

Open access publishing facilitated by University of Wollongong, as part of the Wiley - University of Wollongong agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest to report.

FUNDING INFORMATION

This research did not receive any specific grant from agencies in the public, commercial or not-for-profit sectors.

DATA AVAILABILITY STATEMENT

Data are available upon reasonable request of the authors.

REFERENCES

1. Palmer JD, Sebastian NT, Chu J, et al. Single-isocenter multitarget stereotactic radiosurgery is safe and effective in the treatment of multiple brain metastases. *Adv Radiat Oncol*. 2020;5(1):70-76. doi:10.1016/j.adro.2019.08.013
2. Alongi F, Fiorentino A, Gregucci F, et al. First experience and clinical results using a new non-coplanar mono-isocenter technique (HyperArc™) for Linac-based VMAT radiosurgery in brain metastases. *J Cancer Res Clin Oncol*. 2019;145(1):193-200. doi:10.1007/s00432-018-2781-7
3. Ruggieri R, Naccarato S, Mazzola R, et al. Linac-based radiosurgery for multiple brain metastases: comparison between two mono-isocenter techniques with multiple non-coplanar arcs. *Radiother Oncol*. 2019;132:70-78. doi:10.1016/j.radonc.2018.11.014

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4. Pudsey L, Haworth A, White P, et al. Current status of intra-cranial stereotactic radiotherapy and stereotactic radiosurgery in Australia and New Zealand: key considerations from a workshop and surveys. *Phys Eng Sci Med*. 2022;45:251-259. doi:10.1007/s13246-022-01108-4
5. Rose MS, Tirpak L, Van CasterenK, et al. Multi-institution validation of a new high spatial resolution diode array for SRS and SBRT plan pretreatment quality assurance. *Med Phys*. 2020;47(7):3153-3164. doi:10.1002/mp.14153
6. Kim J, Han MC, Lee E, et al. Detailed evaluation of Mobius3D dose calculation accuracy for volumetric-modulated arc therapy. *Phys Med*. 2020;74(April):125-132. doi:10.1016/j.ejmp.2020.05.015
7. Gao J, Liu X. Off-isocenter Winston-Lutz test for stereotactic radiosurgery/stereotactic body radiotherapy. *Int J Med Phys*. 2016;5:154-161.
8. Poder J, Brown R, Porter H, Gupta R, Ralston A. Development of a dedicated phantom for multi-target single-isocentre stereotactic radiosurgery end to end testing. *J Appl Clin Med Phys*. 2018;19(6):99-108. doi:10.1002/acm2.12452
9. Jhaveri J, Chowdhary M, Zhang X, et al. Does size matter? Investigating the optimal planning target volume margin for post-operative stereotactic radiosurgery to resected brain metastases. *J Neurosurg*. 2019;130:797-803. doi:10.3171/2017.9.JNS171735
10. Hardcastle N, WA Tome. On a single isocenter volumetric modulated arc therapy SRS planning technique for multiple brain metastases. *J Radiosurg SBRT*. 2012;2:1-9.
11. Meeks SL, Mercado CE, Popple RA, et al. Practical considerations for single isocenter LINAC radiosurgery of multiple brain metastases. *Pract Radiat Oncol*. 2022;12(3):195-199. doi:10.1016/j.prro.2021.09.007
12. Pudsey LMM, Biasi G, Ralston A, Rosenfeld A, Poder J. Detection of rotational errors in single-isocenter multiple-target radiosurgery: is a routine off-axis Winston–Lutz test necessary? *J Appl Clin Med Phys*. 2022;23:e13665. doi:10.1002/acm2.13665
13. Agnew A, Agnew CE, Grattan MWD, Hounsell AR, McGarry CK. Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries. *Phys Med Biol*. 2014;59(9):N49-N63. doi:10.1088/0031-9155/59/9/N49
14. Neal B, Ahmed M, Kathuria K, Watkins T, Wijesooriya K, Siebers J. A clinically observed discrepancy between image-based and log-based MLC positions. *Med Phys*. 2016;43(6):2933-2935. doi:10.1118/1.4949002
15. Glenn MC, Brooks FE, Peterson CB, et al. Photon beam modeling variations predict errors in IMRT dosimetry audits. *Radiother Oncol*. 2022;166:8-14. doi:10.1016/j.radonc.2021.10.021
16. Hardcastle N, Nelms B, O'Connor L, et al. OC-0521 SRS plan quality with variation in modality: results of an international planning competition. *Radiother Oncol*. 2019;133:S272-S273. doi:10.1016/s0167-8140(19)30941-7
17. Glenn MC, Peterson CB, Followill DS, Howell RM, Pollard-Larkin JM, Kry SF. Reference dataset of users' photon beam modeling parameters for the Eclipse, Pinnacle, and RayStation treatment planning systems. *Med Phys*. 2020;47(1):282-288. doi:10.1002/mp.13892
18. Lee YC, Kim Y. Commissioning of the TrueBeam STx 6 MV FFF beam in the RayStation treatment planning system for SRS and SBRT treatments. *Int J Med Physics, Clin Eng Radiat Oncol*. 2021;10(01):16-37. doi:10.4236/ijmpcero.2021.101003
19. Lehmann J, Hussein M, Barry MA, et al. SEAFARER—a new concept for validating radiotherapy patient specific QA for clinical trials and clinical practice: SEAFARER—sensitivity of patient specific quality assurance, PSQA. *Radiother Oncol*. 2022;171:121-128. doi:10.1016/j.radonc.2022.04.019
20. Rangel A, Dunscombe P. Tolerances on MLC leaf position accuracy for IMRT delivery with a dynamic MLC. *Med Phys*. 2009;36(7):3304-3309. doi:10.1118/1.3134244
21. Wilke L, Andratschke N, Blanck O, et al. ICRU report 91 on prescribing, recording, and reporting of stereotactic treatments with small photon beams. *Strahlentherapie und Onkol*. 2019;195(3):193-198.
22. Hernandez V, Saez J, Pasler M, Jurado-Bruggeman D, Jornet N. Comparison of complexity metrics for multi-institutional evaluations of treatment plans in radiotherapy. *Phys Imaging Radiat Oncol*. 2018;5(February):37-43. doi:10.1016/j.phro.2018.02.002
23. 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. doi:10.1002/mp.14396
24. Hernandez V, Hansen CR, Widesott L, et al. What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans. *Radiother Oncol*. 2020;153:26-33. doi:10.1016/j.radonc.2020.09.038

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: May L, Hardcastle N, Hernandez V, Saez J, Rosenfeld A, Poder J. Multi-institutional investigation into the robustness of intra-cranial multi-target stereotactic radiosurgery plans to delivery errors. *Med Phys*. 2024;51:910–921.
https://doi.org/10.1002/mp.16907

24734209, 2024, 2, Downloaded from https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.16907 by Readcube (Labtiva Inc.), Wiley Online Library on [13/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License