Practical Radiation Oncology (2018) **xx**, xxx–xxx

practical radiation oncology
pro
www.practicalradonc.org

Basic Original Report

Stereotactic radiosurgery for benign brain tumors: Results of multicenter benchmark planning studies

David J. Eaton PhD [a,*], Jonathan Lee MSc [a], Rushil Patel MSc [a], Antony E. Millin PhD [b], Ian Paddick MSc [c, d], Christopher Walker BSc [e]

[a]*National Radiotherapy Trials Quality Assurance group, Mount Vernon Hospital, Northwood, United Kingdom*
[b]*Velindre Cancer Centre, Cardiff, United Kingdom*
[c]*Medical Physics Ltd, Reading, United Kingdom*
[d]*Cromwell Hospital, London, United Kingdom*
[e]*Northern Centre for Cancer Care, Newcastle, United Kingdom*

Received 22 December 2017; revised 30 January 2018; accepted 13 February 2018

Abstract

Purpose: Stereotactic radiosurgery (SRS) is strongly indicated for treatment of surgically inaccessible benign brain tumors. Various treatment platforms are available, but few comparisons have included multiple centers. As part of a national commissioning program, benchmark planning cases were completed by all clinical centers in the region.

Methods and materials: Four benign cases were provided, with images and structures predelineated, including intracanalicular vestibular schwannoma (VS), larger VS, skull base meningioma, and secreting pituitary adenoma. Centers were asked to follow their local practice, and plans were reviewed centrally using metrics for target coverage, selectivity, gradient falloff, and normal tissue sparing.

Results: Sixty-eight plans were submitted using 18 different treatment platforms. Fourteen plans were subsequently revised following feedback, and review of 5 plans led to a restriction of service on 2 platforms (2 centers). Prescription doses were consistent for VS and meningioma submissions, but a wide range of doses were used for the pituitary case. All centers prioritized coverage, with the prescription isodose covering ≥95% of 78/82 target volumes. Lower values may be expected next to air cavities when using advanced algorithms, and in general may be acceptable for some benign lesions. Selectivity was much more variable, and in some cases this was combined with high gradient index and/or >1 mm margin, resulting in large volumes of normal tissue being irradiated. Normal tissue doses were more variable across linear accelerator (LINAC)-based plans than with Gamma Knife or CyberKnife, and dose spillage seemed independent of prescription isodose

Part of this work was presented at the 13th Congress of the International Stereotactic Radiosurgery Society, Montreux, Switzerland, May 31, 2017.

Sources of support: The Stereotactic Radiosurgery commissioning program was delivered and funded by NHS England, and the Radiotherapy Trials Quality Assurance group is funded by the National Institute for Health Research.

Conflict of interest. I.P. is a consultant for Elekta Instrument AB.

* Corresponding author. Radiotherapy Physics, Mount Vernon Hospital, Northwood HA62RN, United Kingdom.
E-mail address: davideaton@nhs.net (D.J. Eaton).

https://doi.org/10.1016/j.prro.2018.02.006

(inhomogeneity). This may reflect the variety of LINAC-based approaches represented or the necessary tradeoff between different objectives.
Conclusions: These benchmarking exercises have highlighted areas of different clinical practice and priorities and potential for improvement. The subsequent sharing of plan data and margin philosophies between the neurosurgery and oncology communities allowed for meaningful comparison between centers and their peers.

Introduction

Stereotactic radiosurgery (SRS) typically involves highly conformal dose distributions, given with high positional accuracy in 1 fraction. SRS is strongly indicated for treatment of surgically inaccessible benign brain tumors.[1-3] Various treatment platforms are now available, potentially allowing expansion of services beyond specialist units for easier patient access. Comparisons of plan quality have been performed between different platforms, either in single center studies[4,5] or a few multicenter studies.[6-8] Typically, these have included both linear accelerator (LINAC)-based techniques, such as dynamic conformal arcs (DCAs) and volumetric modulated arc therapy (VMAT), along with dedicated units such as Gamma Knife (GK) and CyberKnife (CK), which use multiple noncoplanar beams. It is not clear, however, how well these comparisons represent current clinical practice across a wide region.

Variations in planning and treatment quality are highly undesirable in clinical trials because they can cause significant variation in outcomes, potentially undermining the conclusions of the study.[9] National and international quality assurance (QA) bodies have been established to minimize such variation.[10] Robust programs typically include dosimetry audits and benchmark cases, which are standard cases contoured or planned by participating centers and then analyzed centrally and compared with other submissions.

In 2016, a prerequisite for all providers selected as SRS/stereotactic radiation therapy centers in England was participation in a QA process, informed through collaboration between the national trials QA group and a multidisciplinary expert advisory group consisting of clinical oncologists, neurosurgeons, and physicists. All clinical centers undertook planning benchmark cases, providing a unique dataset of current practice across a large number of providers and a wide range of equipment. Centers were asked to follow their local practice, rather than providing any specific guidance such as a trial protocol. Results were used to assess the technical competency of each successful bidder and to facilitate sharing of best practice, identify outliers, and support centers with less experience. Metastases planning cases have been reported separately.[11]

Methods

Four benign planning cases were distributed to each provider, as previously described,[11] including computed tomography images, predrawn Digital Imaging and Communications in Medicine-radiation therapy (DICOM-RT) structure sets, and a brief clinical history, as follows.

- Case 1: 52-year-old pilot presented with mild tinnitus. Hearing good to excellent: Gardner-Robertson grade 1. Radiological diagnosis of (right intracanalicular) vestibular schwannoma (VS) and proven growth over 1 year. Preservation of hearing vital for the patient's employment. Volumes provided included gross target volume (GTV) (0.07 mL), cochlea, brainstem, trigeminal nerve, and brain.
- Case 2: Fit 70-year-old woman presented with progressive unilateral hearing loss: Gardner-Robinson grade 4 (poor) when referred. Facial function normal: House-Brackmann grade 1. Radiological diagnosis of VS with proven growth >1 year. Volumes provided included GTV (1.9 mL), cochlea, brainstem, trigeminal nerve, pituitary fossa, and normal brain (brain – GTV).
- Case 3: 46-year-old woman presented with increasing facial numbness. Radiological diagnosis of skull base meningioma and proven growth >6 months. Volumes provided included GTV (1.6 mL), cochlea, brainstem, trigeminal nerve, and normal brain.
- Case 4: 36 year old with Cushing syndrome. Transsphenoidal surgical resection 3 years previously; tumor histology was adrenocorticotropic hormone positive pituitary adenoma. Now recurred and patient not keen for another operation and so referred for SRS. Volumes provided included GTV (1.1 mL), brainstem, optic chiasm, optic nerves, pituitary stalk, and normal brain.

Centers produced treatment plans for some or all of the lesions according to their own clinical practice. They were asked not to modify any provided structures, but to add a planning target volume (PTV) margin if this was local practice. Various parameters including prescription doses, margins used, volumes, and dose-volume parameters to

targets and organs at risk (OARs) were recorded, as calculated on the local treatment planning system (TPS). They were used as the primary dataset for central analysis. Structures, plan, and dose cube data were also returned in DICOM-RT format. Review software VODCA (version 5.41, Medical Software Solutions GmbH, Hagedorn, Switzerland) was used as an independent check of submitted values and to correct any gross errors or omissions.

The following plan quality metrics for conformity or dose falloff were calculated, as recommended by international consensus.[12] The first 3 measure the degree of under- or overcoverage of the target by the prescription isodose, with larger values corresponding to better conformity, up to a maximum of 1. The final 2 assess the dose falloff into the surrounding tissue, a key hallmark of SRS, with lower values being optimal. Half prescription isodose (R50%) is typically used for lung stereotactic body radiation therapy plans and combines the effect of selectivity and dose gradient.[13]

Target coverage ratio	=	$\frac{\text{PTV V100\% (mL, TTV)}}{\text{PTV (mL)}}$
Selectivity index	=	$\frac{\text{PTV V100\% (mL, TTV)}}{\text{Total V100\% (mL, PIV)}}$
PCI[14]	=	Coverage * selectivity
GI[15]	=	$\frac{\text{Total V50\% (mL)}}{\text{Total V100\% (mL, PIV)}}$
R50%[13]	=	$\frac{\text{Total V50\% (mL)}}{\text{PTV (mL)}}$

V50% or V100% is the volume of a region receiving 50% or 100% of the prescription dose, respectively. Total V100% is commonly known as the prescription isodose volume (PIV) and PTV V100% is known as treated target volume (TTV). Where no PTV margin was added, GTV volumes were used directly. PCI indicates Paddick conformity index; GI indicates gradient index.

Results

A total of 68 plans were submitted initially using 18 different treatment platforms in 18 centers, as shown in Table 1. The majority was considered acceptable by the multidisciplinary expert advisory group, but 14 plans were subsequently revised following feedback. Review of 5 plans led to a restriction of service in 2 centers, with cases referred to another center in the region.

Prescription doses were consistent for cases 1 and 2 (VS) and case 3 (meningioma) submissions, typically 12 to 13 Gy and 12 to 15 Gy in 1 fraction, respectively (or 27 Gy in 3 fractions in 2 centers). For case 4 (pituitary), however, a wide range of doses were used: 25 to 30 Gy (GK), 22 to 24 Gy (CK), and 12 to 25 Gy (LINAC). Two centers prescribed <18 Gy, but, after feedback, these were either revised higher or the center agreed to refer these patients elsewhere.

All centers appeared to prioritize coverage, with the prescription isodose covering ≥95% of 78 of 82 target volumes. In 1 submission for case 4 (pituitary), the coverage was only 81%, but this was partly a result of using the collapsed cone algorithm. Unlike the TMR10, ray tracing, or pencil beam algorithms used by most centers, this correctly models lack of lateral scatter close to air cavities and so is a more accurate representation of the delivered dose. In this case, however, the center had also

Table 1 Equipment used for benchmark case submissions, with numbers of platforms shown

Platform		TPS (version)		Algorithm name	Technique	Collimation	PTV margin (meningioma)	Revised
Gamma Knife	7	GammaPlan (10.1, 11.0)	7	TMR10	Multiple noncoplanar beams	Cones	0 mm	-
CyberKnife	3	Multiplan (5.21)	3	Ray tracing	Multiple noncoplanar beams	Cones	0 mm (2), 1 mm(2)	-
Varian LINAC (Novalis/ STx/2100)	6	Eclipse (11.0)	1	AAA (1)	Coplanar/ noncoplanar VMAT	2.5 mm MLC	2 mm	1 mm
		iPlan (4.51-4.54)	4	Pencil beam	Static conformal (2 [a]), DCA (2 [a]), fixed cone arcs (2 [a])	2.5 mm MLC (3), cones (2 [a])	0 mm (3), 1 mm (1)	1 mm (1)
		Pinnacle (9.8)	1	Collapsed cone	Static conformal	2.5 mm MLC	0 mm	-
Elekta LINAC (Synergy/ Agility)	2	Monaco (5.2)	1	Monte Carlo	Noncoplanar VMAT	5 mm MLC	2 mm	1 mm
		Pinnacle (9.6)	1	Collapsed cone	Static conformal	4 mm MLC	2 mm	Not revised

DCA, dynamic conformal arcs; LINAC, linear accelerator; MLC, multileaf collimator; PTV, planning target volume; TPS, treatment planning system; VMAT, volumetric modulated arc therapy; VS, vestibular schwannoma.

[a] One center used fixed cone arcs for case 1 (intracanalicular VS), but static conformal beams for the other 3 cases. Another center initially used DCA for cases 1 and 2 (VS), but static conformal beams for cases 3 and 4 (meningioma and pituitary) and then changed to DCA for case 3 revised submissions.

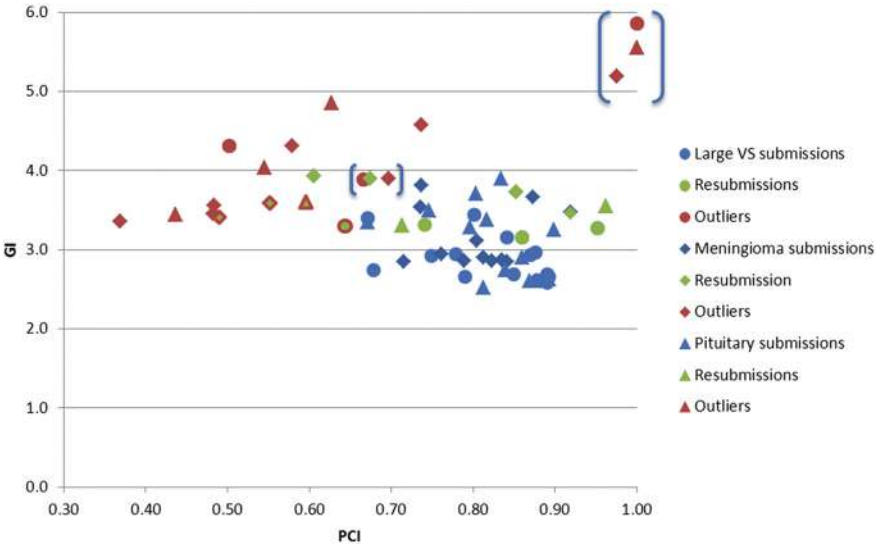

Figure 1 Variation of GI and PCI of submissions for larger volume cases 2 through 4 showing those considered outliers by the expert reference group, as well as resubmissions. Resubmissions that were still deemed outliers are shown in green with a red outline. Five plans in brackets were outliers for other reasons: low coverage (0.87), large PTV margin (2 mm), and 3 cases (top right) that used a large dose grid (2.5-3.0 mm); therefore, reported values may be inaccurate. GI, gradient index; PCI, Paddick conformity index; PTV, planning target volume; VS, vestibular schwannoma.

applied a 1.5 mm PTV margin, extending the PTV into air and thereby exacerbating the effect.

Selectivity (and PCI) was much more variable, and several cases with low values and/or high GI resulted in large volumes of normal tissue being irradiated. Figure 1 shows that outliers typically had PCI <0.65 and/or GI >4.0, and there was also an optimal cluster of plans with PCI 0.75 to 0.90 and GI 2.5 to 3.0. Any further increase in PCI was only possible by a tradeoff of higher GI, and 2.5 seems to reflect an absolute lower limit for GI; however, improvements were possible for resubmissions without changing equipment in several cases, partly based on the feedback on what was achievable in other centers. Table 2 shows PCI and GI values collated by treatment platform or technique, omitting those plans that were revised or led to a restriction of service (ie, the outliers).

The 14 resubmitted plans included the following changes.

Table 2 Conformity and gradient indices reported by this series (omitting outliers that were replanned or those who would not treat) and comparative values from other studies

Platform	Technique	Case mix	PCI	GI	Reference
Gamma Knife		6 VS	0.81 ± 0.03	2.7 ± 0.2	Abacioglu et al[5]
		6 M	0.86 ± 0.05	2.6 ± 0.1	
		10 AVM, 5 VS	0.77 ± 0.04	2.6 ± 0.1	Gevaert et al[7]
		10 M	0.77 ± 0.06	2.7 ± 0.2	Kaul et al[8]
		VS, M, P	0.84 ± 0.05	2.7 ± 0.1	This study
CyberKnife		10 AVM, 5 VS	0.77 ± 0.06	3.5 ± 0.5	Gevaert et al[7]
		10 M	0.76 ± 0.07	3.4 ± 0.1	Kaul et al[8]
		VS, M, P	0.82 ± 0.05	3.3 ± 0.5	This study
Varian LINAC	DCA	10 AVM, 5 VS	0.66 ± 0.04	3.2 ± 0.6	Gevaert et al[7]
	DCA/other [a]	VS, M, P	0.70 ± 0.10	3.4 ± 0.4	This study
	IMRT	10 AVM, 5 VS	0.68 ± 0.04	3.9 ± 0.9	Gevaert et al[7]
		10 M	0.66 ± 0.07	3.5 ± 0.9	Kaul et al[8]
	VMAT	6 VS	0.84 ± 0.04	3.8 ± 0.6	Abacioglu et al[5]
		6 M	0.88 ± 0.05	3.8 ± 0.5	
Varian/Elekta LINAC	VMAT	VS, M, P	0.89 ± 0.06	3.5 ± 0.2	This study

Values are given as mean and 1 standard deviation.

AVM, arteriovenous malformation; DCA, dynamic conformal arcs; GI, gradient index; IMRT, intensity modulated radiation therapy (fixed gantry angle); M, meningioma; P, pituitary adenoma; PCI, Paddick conformity index; VS, vestibular schwannoma. Other abbreviation as in Table 1.

[a] DCA, static conformal, fixed cone arcs. See Table 1 for full breakdown.

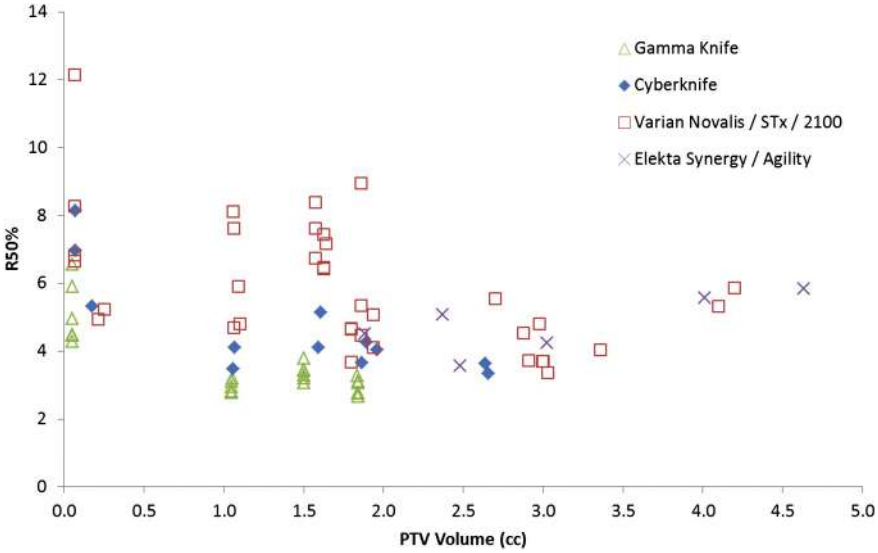

Figure 2 R50% against PTV volume (logarithmic scale) for all lesions, showing distribution for each platform. PTV volumes reflect the margin applied so include the per-platform variation listed in Table 1. R50%, relative spread of half prescription isodose; other abbreviation as in Fig 1.

- Technique (1 submission): Static conformal fields changed to DCA
- Substantial change in number or orientation of fields (6): single coplanar VMAT arc changed to 2 noncoplanar VMAT arcs (3); 8 to 15 fixed cone arcs changed to 19 to 24 arcs (3)
- Margin (6): 2 mm reduced to 1 mm for Varian Truebeam STx (3), 2 mm reduced to 1 mm for Elekta Versa HD (1), and 0 mm increased to 1 mm for Varian Novalis (2)
- Reduction of dose grid (3): 2.5 to 3.0 mm to 1.0 mm isotropic
- Substantial change in prescription isodose (4): 89% to 71% (1), 79% to 50% (1), 75% to 65% (1), 50% to 75% (1)
- Reoptimization of plan alone (1)

Figure 2 shows the R50% against PTV. All platforms were able to give plans with relatively low R50% for larger lesion volumes, but for smaller volumes there was increasing variation both inter- and intratreatment platform. Even for lesions >1 mL, there was more variation among Varian LINAC plans (range, 3.4-8.9) and Elekta LINAC plans (3.6-5.8, although this includes a small

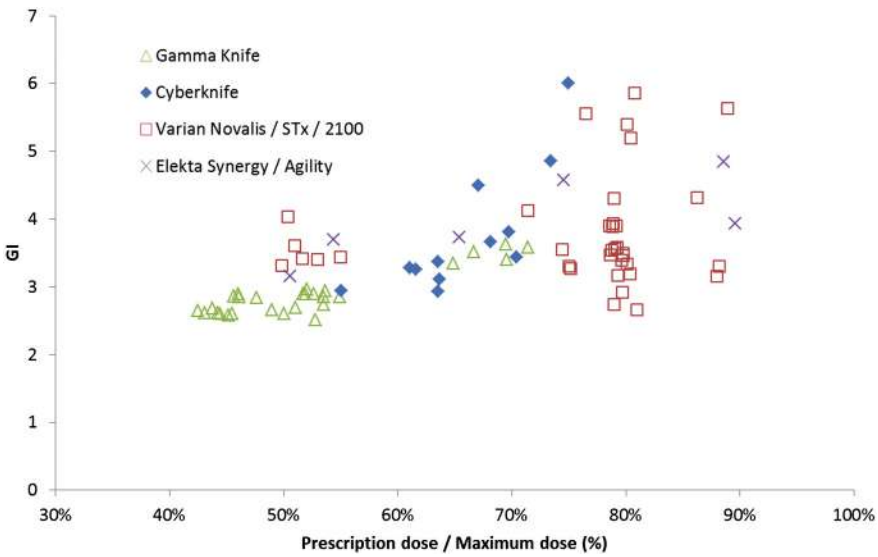

Figure 3 Variation of GI with the prescription isodose level, showing distribution for each platform. Maximum dose was as reported by the treatment planning system. Abbreviation as in Fig 1.

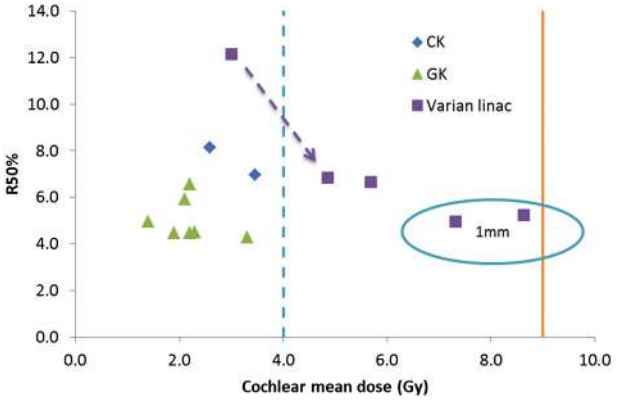

Figure 4 Variation of R50% with cochlea mean dose for case 1 (intracanalicular VS). Two centers using 1 mm PTV margin are circled; the 1 resubmission is indicated by the dashed arrow. Vertical lines correspond to an optimal constraint of 4 Gy and the limit of 9 Gy given in AAPM TG101.[20] AAPM TG101, American Association of Physicists in Medicine Task Group 101; CK, CyberKnife; GK, Gamma Knife. Other abbreviations as in Figs 1 and 2.

number of samples), than for CK and GK (range, 3.3-5.1 and 2.7-3.8, respectively). Figure 3 shows the relationship between GI and prescription isodose level (ie, prescription dose as a percentage of maximum dose, as commonly reported in SRS).

Normal tissue dose variations with respect to R50% are shown for cochlea (case 1, Fig 4), brainstem (cases 2-4, Fig 5), and optic chiasm (case 4, Fig 6). Doses to other OARs were typically small or not considered clinically significant. Cochlea mean doses for case 2 (larger VS, no hearing to preserve) were 3.1 to 12.0 Gy. Optic nerve maximum doses for case 4 (pituitary) were similar to chiasm doses, with mean 4.5 Gy (range, 1.4-9.0). Trigeminal nerve maximum doses were 9.7 to 14.6 Gy (case 2, VS) and 11.1 to 16.9 Gy (case 3, meningioma). Pituitary stalk maximum doses were 6.7 to 18.9 Gy (case 4, pituitary).

Discussion

A wide variety of systems, techniques, and planning parameters were observed across the 4 cases. Prescription doses were mostly similar, except for the pituitary case. A review by Minniti et al[3] recommended doses of ≥12 Gy for nonfunctioning adenomas to ensure high tumor control rates. For Cushing disease (adrenocorticotropic hormone–secreting adenoma such as case 4), they report similar levels of biochemical remission for doses 15 to 20 Gy, 20 to 25 Gy, and >25 Gy, but note that 1 study showed much higher remission rates using greater than 25 Gy; therefore, the optimal dose is yet to be determined. There was a systematic variation in prescribed dose based on equipment, which may reflect the ability of different platforms to spare adjacent normal tissues and escalate the dose.

Target margins

The planning philosophy of all radiation therapy is a high dose to the target and minimal dose to normal tissue. In conventional radiation therapy, the clinical target

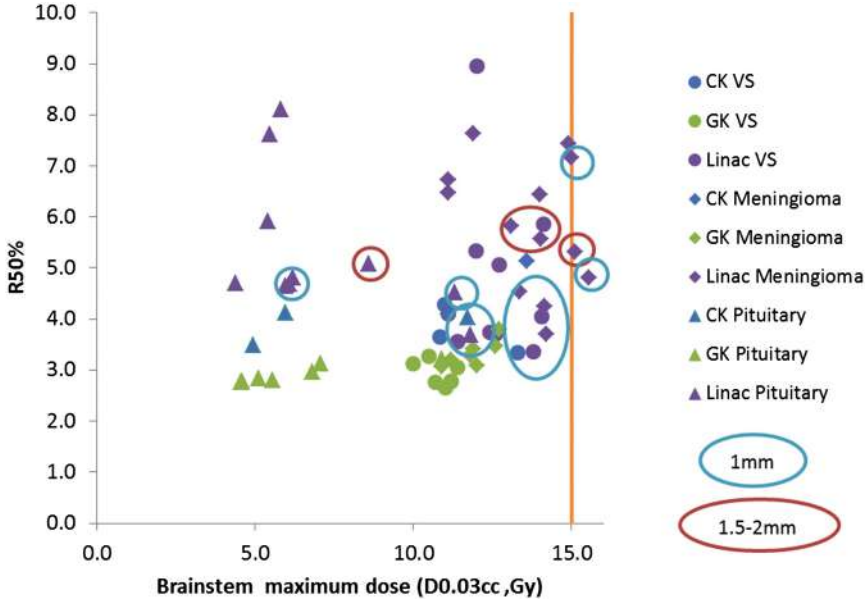

Figure 5 Variation of R50% with brainstem maximum dose (to 0.03 mL) for cases 2 through 4 (larger VS, meningioma, and pituitary). LINAC points include both Varian and Elekta LINAC submissions. Centers using 1 and 1.5 to 2.0 mm PTV margins are circled. The vertical line corresponds to the limit of 15 Gy in AAPM TG101.[20] LINAC, linear accelerator. Other abbreviations as in Figs 1, 2, and 4.

volume usually includes a significant amount of normal tissue; therefore, a homogeneous dose throughout the target is preferable. The SRS philosophy of treating the GTV only (knowing that any inaccuracies will leave the target receiving less than the full prescription dose) parallels the lack of surgical necessity to remove every tumor cell of a benign tumor; tumor rests are commonly left in microsurgery and often do not grow.

For intracranial SRS, the main normal tissue is the brain itself, which is often completely surrounding the target. To minimize the dose to this OAR, a steep dose gradient is placed on the edge of the PTV. This can best be achieved by having a steep dose gradient both on the inside and outside of the PTV. By including a margin around the clinical target volume, a much higher dose is delivered to the target, but also a higher dose to the part of the brain that is now included within the PTV. For example, adding a 1 mm margin to an 8 mm diameter target almost doubles the irradiated volume and will normally mean that more brain is being irradiated within the prescription isodose than tumor.

The fixed geometry of GK ensures spatial uncertainties are minimized and therefore lends itself to a zero margin PTV philosophy. Although uncertainties associated with image distortion, spatial accuracy of the planning imaging system, and transfer errors between these imaging systems and the planning system exist in all platforms, they are a smaller concern than machine geometry issues associated with rotating gantry systems. In many cases, imaging and transfer errors were minimized by use of a stereotactic frame for the planning scan. Even with frameless immobilization, spatial uncertainties were reduced by the use of room-based imaging and a 6 degree of freedom couch, correcting any minor rotations and couch walkout from the isocenter before each treatment field.

Gantry sag, the deviation of the radiation isocenter from the assumed isocenter as the gantry rotates, cannot be accounted for using these systems however and can be a significant contribution to the overall geometric uncertainty of LINAC-based systems. Reported magnitudes of gantry sag were in the 0.2 to 1.7 mm range. In the centers in which 2 mm margins were proposed, feedback suggested enhanced QA and setup of the accelerator to minimize and accurately quantify machine-based uncertainties, so that PTV margins were reduced to 1 mm. For certain benign lesions, some centers changed their practice to add no margin and accept that their positional uncertainties may mean a slightly reduced coverage in practice.

Target conformity

Almost all submissions had >95% coverage; therefore, the PCI values mainly reflect the variation in selectivity (overcoverage). Selectivity <0.5 means that more normal tissue is being irradiated within the prescription dose than target. High values of GI also led to large volumes of normal tissue receiving medium-to-high doses, which is also undesirable. Figure 1 shows that most outliers had PCI <0.65 and GI >4.0, falling in the top and left halves of the graph; these values could be used as a guide for other cases. There further appears to be a cluster of "optimal" submissions with PCI >0.75 and GI <3.0 and an apparent "ideal" limit of about PCI 0.90 and GI 2.5 in the bottom right of the graph. Closer conformity may only be possible by compromising dose falloff and reflects a tradeoff between these 2 parameters.

Suggested optimal values for GK treatments are PCI >0.85 and GI <3.0.[12,15] One study found worse toxicity with GI <3.0 for meningioma cases, but this effect disappeared when testing the data with multivariate analysis.[4] Other plan comparison studies are typically limited to a single center per platform, but similar trends may be observed, as shown in Table 2. GK, CK, and VMAT plans are able to achieve high values of PCI compared with other LINAC-based techniques; however, only GK plans were able to consistently achieve low GI values. Other authors have noted that GK treatments can take longer to deliver than other modalities,[6-8] especially LINAC-based VMAT delivery, which may also reduce the radiobiological effective dose.[16]

There are insufficient data to show whether multileaf collimator size is a factor in plan quality or if there are systematic differences between Varian and Elekta LINACs or different TPSs for these platforms (Table 2, Fig 2). One other study suggested that multileaf collimator -based LINACs could be used as in addition to specialist systems for lesions >0.5 mL,[6] but used an older GK model and applied 2 mm margins to all systems, which is not representative of clinical practice. Figure 2, however, shows that several centers were able to achieve low values of R50%, which includes the effect of both selectivity and gradient index. As PTV volume decreases, it is expected that R50% will increase; however, both Varian and Elekta LINAC plans showed greater variation for a given volume, which may reflect the variety of equipment and techniques within these categories, as well as some systematic differences in prescription doses and margins applied.

Other authors have noted that LINAC-based plans are typically more homogeneous within the target,[7] but this is not an objective for SRS and may well have led to the poorer GI values in these comparisons. Figure 3 shows that for GK and CK plans, there appears to be an increasingly steep rise in GI with prescription isodose (as defined relative to the maximum dose). For other LINAC-based plans, no such trend is apparent. In general, the different modalities use different ranges, similar to a national survey of SRS practice that reported: 45% to 55% for GK, 55% to 80% for CK, and 80% to 100% for other LINACs plans.[17] Some LINAC-based plans, however, used low isodoses of about 50%, but with higher GI than GK or CK plans at the same level. Conversely, many Varian and Elekta LINACs plans prescribed to the 80% isodose, with considerable variation in GI. Some of these were able to achieve low GI

values, although about one-half of these cases were also associated with lower selectivity.

Normal tissue doses

Normal tissue doses were more variable across Varian and Elekta LINACs plans than GK or CK, which may reflect the variety of approaches represented or the necessary tradeoff between different objectives with these platforms. In general, use of PTV margins led to higher OAR doses as expected, because the target volume will be closer when these expansions are used. One other planning study found OAR doses were similar between GK and VMAT plans,[5] but this was limited to a single center, and there are systematic differences apparent in our data.

There are no widely accepted consensus tolerances for normal tissues during SRS, partly because of variation in contouring and reporting practice.[12,18] Some centers may use guidelines for stereotactic body radiation therapy by the American Association of Physicists in Medicine Task Group 101 (TG101),[19] but more recent clinical data are also available.[20]

For cochlea, TG101 suggests a near-point-maximum dose constraint of 9Gy (20), but the cochlea is often so small that mean dose is a better surrogate than near-point maximum dose (to 0.035 mL). All centers achieved a mean dose <9 Gy for case 1 (intracanalicular VS), where hearing preservation was required (Fig 4). Recently, several studies have suggested that mean doses less than 4.0 to 5.3 Gy increase the likelihood of hearing preservation ,[19-22] so a 4 Gy mean dose may be a better optimal constraint.[21] Only GK, CK and 1 LINAC submission for case 1 were able to meet this limit, but this may reflect the PTV margins added. In this situation, when the OAR is in very close proximity to the target on 1 side, there can be a tradeoff between sparing of dose in that direction (to reduce OAR dose) and limiting spread of dose in all directions (to reduce GI or R50%).

For brainstem, maximum doses of 12 to 14 Gy have been associated with low risk of neurological complications,[3] and TG101 has a near-point-maximum constraint of 15 Gy.[20] Figure 5 shows that almost all centers kept within this constraint, but some cases were more challenging than others. For case 4 (pituitary), some centers kept brainstem doses low, but at the expense of higher R50% (medium dose spread in other directions), whereas some allowed higher brainstem doses to reduce R50% overall. For case 3 (meningioma), it was more challenging to achieve low brainstem doses for all submissions, so the potential for this tradeoff was more limited. In general, submissions with R50% <5 were typically considered acceptable, and all platforms types (but not all centers) were able to achieve these levels.

For optic apparatus (including chiasm and optic nerves), maximum tolerance doses between and 8 and 12 Gy have been suggested, to limit the risk of optic neuropathy.[3,12,19] In this study, all submissions for case 4 (pituitary, Fig 6) had near-point-maximum doses <10 Gy, in line with TG101[20]; however, it may be that for some submissions prescription doses were limited to satisfy this dose constraint. A similar tradeoff is seen in this case between OAR dose and R50%, especially for LINAC-based plans, and the largest doses typically correspond to centers who have applied a PTV margin. For trigeminal nerve, no typical constraints are available and it is likely that centers did not explicitly spare this OAR in many cases.

Impact and limitations

This study is the largest known planning evaluation of different SRS centers and platforms, but has some limitations. Because only a limited number of cases were used for practical reasons, they may not have reflected the typical case mix seen by different providers. Because no

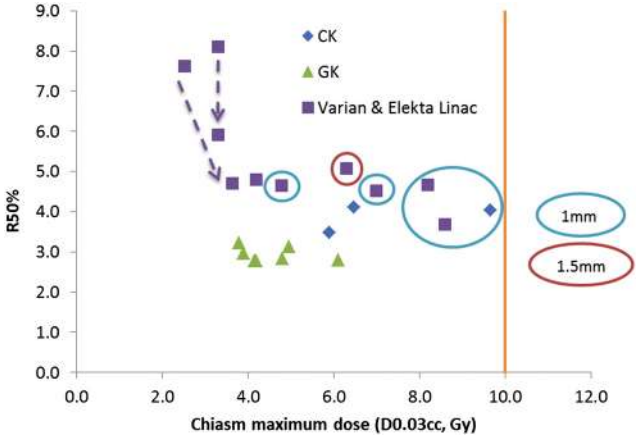

Figure 6 Variation of R50% with optic chiasm maximum dose (to 0.03 mL) for case 4 (pituitary). Centers using 1 and 1.5 mm PTV margins are circled; the 2 resubmissions are indicated by the dashed arrows. The vertical line corresponds to the limit of 10 Gy in AAPM TG101.[20] Abbreviations as Figs 1, 2, and 4.

guidance was provided to participating centers, the results reflect the current clinical practice of a particular geographic region and systems in their native environment. Previous platform studies were often based on preset goals in terms of planning metrics, and this can prejudice measurement of the actual performance of particular platforms. Measuring what is delivered is more important than measuring what could potentially be achieved if peculiar parameters are the only objective.

We did not collect information about the time taken to produce treatment plans (or deliver them), so it not known whether centers produced more complex plans than would be practical. For example, adding additional isocenters to a GK plan, beams to a CK plan, or additional arcs to a LINAC-based plan can improve the conformity, but at the expense of long delivery times; however, acceptable plans typically used similar numbers of beams: VMAT 2 to 5 arcs, DCA 3 to 6 arcs, and static conformal plans 7 to 10 beams (although 1 center used 19).

Every attempt was made to minimize corruption of data during transfer between systems, but it is recognized that different methods of volume and dose calculation will lead to variation in reported values. Differences in the calculation of the same planning parameters between TPS may give variations of 5% to 10%,[23,24] but these are unlikely to change the overall trends and conclusions in this study. In addition, individual planning metrics should always be considered in context when determining the acceptability of a specific treatment plan.

Value of process for achieving service improvement

This process was highly unusual in terms of a QA process in that a benchmark was achieved based on the centers rather than current practices. The subsequent sharing of plan data and margin philosophies between the neurosurgery and oncology communities allowed for meaningful comparison between centers and their peers. This open approach encouraged centers to work with partner centers with similar equipment to improve their submissions. Where equipment limitations did not allow for sufficient improvement in plan quality, the subsequent service restrictions were readily evidenced and accepted. This approach has given assurance of the safety and quality of SRS delivery on a national scale and may be useful for future optimization of other treatment sites and techniques.

Conclusion

A national benchmarking exercise for SRS planning has highlighted some variation in clinical practice and priorities, with several outliers that have led to revision of local practice or limitations on clinical use. Most platforms were able to achieve acceptable plans, especially after feedback on what was achievable for these cases. Varian and Elekta LINAC plans were more variable, however, and dose spillage seemed independent of prescription isodose used. Only GK plans appeared to avoid tradeoffs between different objectives, such between doses to critical organs (OARs) and general dose spillage (as quantified by GI or R50%). Overall, this approach has given confidence in the safe and consistent delivery of SRS services across multiple centers, and can also provide useful guidance for centers worldwide.

Acknowledgments

The authors are grateful to all those who supported the SRS QA program, including Elizabeth Miles (RTTQA), Kim Fell (NHS England) and the expert advisory group: Steve Bolton, Andrew Brodbelt, Alison Cameron, Phil Cooper, Martyn Gilmore, Katharine Hunt, Caroline Jones, Catherine McBain, Bodo Lippitz, Tony Millin, Ian Paddick, Matthias Radatz, Frank Saran, Chris Walker and Lucy Winch. We are also very grateful to the staff at all the participating centers.

References

1. Santacroce A, Walier M, Régis J, et al. Long-term tumor control of benign intracranial meningiomas after radiosurgery in a series of 4565 patients. *Neurosurgery.* 2011;70:32-39.
2. Lipski SM, Hayashi M, Chernov M, et al. Modern Gamma Knife radiosurgery of vestibular schwannomas: Treatment concept, volumetric tumor response, and functional results. *Neurosurg Rev.* 2015;38:309-318.
3. Minniti G, Osti MF, Niyazi M. Target delineation and optimal radiosurgical dose for pituitary tumors. *Radiat Oncol.* 2016;11:135.
4. Balagamwala EH, Suh JH, Barnett GH, et al. The importance of the conformality, heterogeneity, and gradient indices in evaluating Gamma Knife radiosurgery treatment plans for intracranial meningiomas. *Int J Radiat Oncol Biol Phys.* 2012;83:1406-1413.
5. Abacioglu U, Ozen Z, Yilmaz M, et al. Critical appraisal of RapidArc radiosurgery with flattening filter free photon beams for benign brain lesions in comparison to Gamma Knife: A treatment planning study. *Radiat Oncol.* 2014;9:119.
6. Schoonbeek A, Monshouwer R, Hanssens P, et al. Intracranial radiosurgery in the Netherlands: A planning comparison of available systems with regard to physical aspects and workload. *Tech Cancer Res Treat.* 2010;9:279-289.
7. Gevaert T, Levivier M, Lacornerie M, et al. Dosimetric comparison of different treatment modalities for stereotactic radiosurgery of arteriovenous malformations and acoustic neuromas. *Radiat Oncol.* 2013;106:192-197.
8. Kaul D, Badakhshi H, Gevaert T, et al. Dosimetric comparison of different treatment modalities for stereotactic radiosurgery of meningioma. *Acta Neurochir.* 2015;157:559-564.
9. Ohri N, Shen X, Dicker AP, et al. Radiotherapy protocol deviations and clinical outcomes: A meta-analysis of Cooperative Group clinical trials. *J Natl Cancer Inst.* 2013;106:387-393.
10. Melidis C, Bosch WR, Izewska J, et al. Global harmonization of quality assurance naming conventions in radiation therapy clinical trials. *Int J Radiat Oncol Biol Phys.* 2014;90:1242-1249.

11. Eaton DJ, Lee J, Paddick I. Stereotactic radiosurgery for multiple brain metastases: Results of multi-centre benchmark planning studies [e-pub ahead of print]. Pract Radiat Oncol https://dx.doi.org/10.1016/j.prro.2017.12.011, accessed February 28, 2018.
12. Torrens M, Chung C, Chung H-T, et al. Standardization of terminology in stereotactic radiosurgery: Report from the Standardization Committee of the International Leksell Gamma Knife Society. *J Neurosurg.* 2014;121(suppl 2):2-15.
13. Timmerman R, Paulus R, Galvin J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA.* 2010;303:1070-1076.
14. Paddick I. A simple scoring ratio to index the conformity of radiosurgical treatment plans. *J Neurosurg.* 2000;93(suppl 3):219-222.
15. Paddick I, Lippitz B. A simple dose gradient measurement tool to complement the conformity index. *J Neurosurg.* 2006;105(Suppl):194-201.
16. Millar WT, Hopewell JW, Paddick I, et al. The role of the concept of biologically effective dose (BED) in treatment planning in radiosurgery. *Phys Med.* 2015;31:627-633.
17. Dimitriadis A, Kirkby KJ, Nisbet A, et al. Current status of cranial stereotactic radiosurgery in the UK. *Br J Radiol.* 2016;89:20150452.
18. Sandström H, Chung C, Jokura H, et al. Assessment of organs-at-risk contouring practices in radiosurgery institutions around the world – The first initiative of the OAR Standardization Working Group. *Radiother Oncol.* 2016;121:180-186.
19. Benedict SH, Yenice KM, Followill D, et al. Stereotactic body radiation therapy: The report of AAPM Task Group 101. *Med Phys.* 2010;37:4078-4101.
20. Hanna GG, Murray L, Patel R, et al. UK consensus on normal tissue dose constraints for stereotactic radiotherapy. *Clin Oncol (R Coll Radiol).* 2018;30:5-14.
21. Tamura M, Carron R, Yomo S, et al. Hearing preservation after Gamma Knife radiosurgery for vestibular schwannomas presenting with high-level hearing. *Neurosurgery.* 2009;64:289-296.
22. Yomo S, Carron R, Thomassin J-M, et al. Longitudinal analysis of hearing before and after radiosurgery for vestibular schwannoma. *J Neurosurg.* 2012;117:877-885.
23. Ma L, Sahgal A, Nie K, et al. Reliability of contour-based volume calculation for radiosurgery. *J Neurosurg.* 2012;117(Suppl):203-210.
24. Eaton DJ, Alty K. Dependence of volume calculation and margin growth accuracy on treatment planning systems for stereotactic radiosurgery. *Br J Radiol.* 2017;90:20170633.