Practical Radiation Oncology

Benchmarking Tests of Contemporary SRS Platforms: Have Technological Developments Resulted in Improved Treatment Plan Quality?

--Manuscript Draft--

Manuscript Number:	PRACTICALRADONC-D-23-00037R1
Article Type:	Basic Original Report
Section/Category:	Physics Contribution
Corresponding Author:	Ian Paddick, BSc(Hons), MSc Streatley, Berkshire UNITED KINGDOM
First Author:	Ian Paddick
Order of Authors:	Ian Paddick
	Judith Mott
	James Bedford
	Petr Filatov
	Diana Grishchuk
	Gavin Orchin
	Peter Houston
	David J Eaton
Abstract:	Introduction The technological evolution of SRS equipment is significant and poses the question “have technological improvements led to a corresponding improvement in treatment plans?”. The following platforms were selected as ‘state of the art’ in 2022: Gamma Knife Icon (GK), CyberKnife S7, Brainlab Elements (Elekta VersaHD and Varian TrueBeam), Varian Edge with HyperArc, Zap-X. Methods and Materials Six benchmarking cases were used from a 2016 study1,2. To reflect the evolution of increasing numbers of metastases treated per patient, a 14-target case was added. The 28 targets amongst the seven patients ranged from 0.02cc to 7.2cc in volume. Participating centres were sent images and contours for each patient and asked to plan to the best of their ability. While some variation in local practice was allowed (e.g. margins), groups were asked to prescribe a specified dose to each target and tolerance doses to organs at risk were agreed. Parameters compared included coverage, selectivity, Paddick Conformity Index (PCI), Gradient Index (GI), R50%, Efficiency Index (EI), doses to OARs, planning and treatment times. Results Mean coverage for all targets ranged from 98.2% (Brainlab/Elekta) to 99.7% (HyperArc 6X). PCI values ranged from 0.722 (Zap-X) to 0.894 (CyberKnife). GI ranged from a mean of 3.52 (GK), representing the steepest dose gradient to 5.08 (HyperArc 10X). The GI appeared to follow a trend with beam energy, with the lowest values from the lower energy platforms (GK; 1.25MeV, Zap-X; 3MV) and the highest value from the highest energy (HyperArc 10X). R50% values had a minimum mean value of 4.48 (GK) and a maximum mean value of 5.98 (Hyperarc 10X). Treatment times were lowest for C-arm linacs. Conclusion Compared with earlier studies, newer equipment appears to deliver higher quality treatments. CyberKnife and linac platforms appear to give better conformity while lower energy platforms give better dose gradient.

Title: Benchmarking Tests of Contemporary SRS Platforms: Have Technological Developments Resulted in Improved Treatment Plan Quality?

Short Running Title: UK SRS Benchmarking

Ian Paddick, MSc, National Hospital for Neurology and Neurosurgery, London, United Kingdom

Judith Mott, PhD, Northern Centre for Cancer Care, Newcastle upon Tyne Hospital NHS Foundation Trust, UK

James Bedford, PhD, Joint Department of Physics, The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London, UK.

Petr Filatov, MSc GenesisCare, Oxford, UK

Diana Grishchuk, MSc, National Hospital for Neurology and Neurosurgery, London, United Kingdom

Gavin Orchin, PhD, Beatson West of Scotland Cancer Centre, NHS Greater Glasgow & Clyde, Glasgow, UK

Peter Houston, MSc, Beatson West of Scotland Cancer Centre, NHS Greater Glasgow & Clyde, Glasgow, UK

David J Eaton, PhD, Department of Medical Physics, Guy's and St Thomas' NHS Foundation Trust, School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK

Corresponding Author Name & Email Address

Ian Paddick (ian@physicsconsulting.co.uk)

Author Responsible for Statistical Analysis Name & Email Address

Diana Grishchuk (diana@physicsconsulting.co.uk)

Conflict of Interest Statement for All Authors

Ian Paddick - Conflict of Interest: works as an ad-hoc consultant for Elekta. Past educational seminars with Elekta AB and Zap Surgical. Board Member: International Stereotactic Radiosurgery Society (ISRS) – serving as Past President

Judith Mott – Conflict of Interest: None

James Bedford – Conflict of Interest: Accuray Inc research agreement with funding to ICR/RMH and research agreement without funding

Petr Filatov - Conflict of Interest: Brainlab

Diana Grishchuk - Conflict of Interest: Brainlab

Gavin Orchin – Conflict of Interest: None

Peter Houston - Conflict of Interest: None

David J Eaton - Conflict of Interest: None

Funding Statement

Funding: None

Data Availability Statement for this Work

Research data are not available

Acknowledgements

We are very grateful to Sakil Zuberi, Honorata Chajecka-Szczygielska, Rekha Mohanraj, Abdul Hussein Mroue for their support in treatment plan creation and extraction of planning parameters, and Christopher Walker for his support in setting up the study.

Abstract

Introduction

SRS treatment delivery can be performed with a range of devices, each of which have evolved over recent years. We sought to evaluate the differences in performance of contemporary SRS platforms and also to compare them with earlier platform iterations from a previous benchmarking study.

Methods and Materials

The following platforms were selected as 'state of the art' in 2022: Gamma Knife Icon (GK), CyberKnife S7 (CK), Brainlab Elements (Elekta VersaHD and Varian TrueBeam), Varian Edge with HyperArc (HA) and Zap-X. Six benchmarking cases were used from a 2016 study[1,2]. To reflect the evolution of increasing numbers of metastases treated per patient, a 14-target case was added. The 28 targets amongst the seven patients ranged from 0.02cc to 7.2cc in volume.

Participating centres were sent images and contours for each patient and asked to plan them to the best of their ability. While some variation in local practice was allowed (e.g. margins), groups were asked to prescribe a specified dose to each target and tolerance doses to organs at risk were agreed.

Parameters compared included coverage, selectivity, Paddick Conformity Index (PCI), Gradient Index (GI), R50%, Efficiency Index (EI), doses to organs at risk (OARs), planning and treatment times.

Results

Mean coverage for all targets ranged from 98.2% (Brainlab/Elekta) to 99.7% (HA-6X). PCI values ranged from 0.722 (Zap-X) to 0.894 (CK). GI ranged from a mean of 3.52 (GK), representing the steepest dose gradient to 5.08 (HA-10X). The GI appeared to follow a trend with beam energy, with the lowest values from the lower energy platforms (GK; 1.25MeV, Zap-X; 3MV) and the highest value from the highest energy (HA-10X). Mean R50% values ranged from 4.48 (GK) to 5.98 (HA-10X). Treatment times were lowest for C-arm linacs.

Conclusion

Compared with earlier studies, newer equipment appears to deliver higher quality treatments. CyberKnife and linac platforms appear to give higher conformity while lower energy platforms yield a steeper dose gradient.

Introduction

There are several competing technologies used for treatment planning and delivery of stereotactic radiosurgery (SRS) and it is important to understand their relative performance. The 2016 UK benchmarking study, which included 24 centres each planning six different SRS plans, was initiated by NHS England and conducted by the Radiotherapy Trials and Quality Assurance Group (RTTQA). This was the largest study of its kind and captured a valuable snapshot of the capabilities and practices of centres commissioned by NHS England to provide intracranial treatments[1,2].

The technological evolution of radiotherapy equipment, particularly in the field of SRS, where the demands of conformity, dose gradient and spatial accuracy are highest, is noteworthy. Six years on, new technology and treatment delivery techniques pose the question “have technological improvements led to a corresponding improvement in treatment

plans?". This paper therefore describes a benchmarking study devised to assess the capabilities of contemporary equipment.

Methods and Materials

Platforms

The following platforms were selected as being 'state of the art' in 2022:

GK with Lightning inverse planning (Elekta Instruments, Sweden), CK S7 with M6 MLC (Accuray, USA), Brainlab Elements (Brainlab, Germany) in conjunction with both Elekta Versa HD (EE) and Varian TrueBeam (EV) (Varian Medical Systems, USA), Varian Edge with HA and Zap-X (Zap Surgical, USA). A multi-centre study was adopted, incorporating six centres, each of which had considerable experience in the platform they represented. Three of these centres had participated in the original benchmarking study. STable1, in Supplemental material 1, details the treatment platforms selected for the study. Due to limitations of the tools available within each treatment planning system to extract planning parameters, special procedures ensured that the metrics were collected as uniformly as possible between the different platforms (Supplemental material 2).

Patient cases and planning objectives

The same six patient plans (two multiple metastasis cases, four benign targets) were used from the previous studies[1,2]. The indications, doses prescribed and treatment planning instructions are summarised below. Patient medical history is described elsewhere[1,2]. To reflect the evolution of SRS on the increased number of metastases treated per patient, a patient with 14 targets was added as a seventh case.

- Case 1: Three cerebral metastases. Prescription doses for PTV 1&2: 24Gy (GK, CK and Zap-X), 21Gy (C-arm linacs), PTV3 (in brainstem): 18Gy (GK, CK and Zap-X), 15Gy (C-arm linacs).
- Case 2: Seven cerebral metastases. Prescription doses for PTV1: 20Gy (GK, CK and Zap-X), 18Gy (C-arm linacs), PTV2-4: 22Gy (GK, CK and Zap-X), 21Gy (C-arm linacs), PTV5-7: 24Gy (GK, CK and Zap-X), 21Gy (C-arm linacs).
- Case 3: Small intracanalicular vestibular schwannoma. Preservation of hearing is vital for the patient to retain his job as a pilot. Prescription dose: 12Gy. Aim for mean dose to cochlea $\leq$ 4Gy.
- Case 4: Vestibular schwannoma. No useful hearing. Prescription dose: 12Gy. Report maximum dose to the brainstem – $D_{0.030cc}$.
- Case 5: Skull base meningioma. Prescription dose: 14Gy. Report maximum dose to the brainstem – $D_{0.030cc}$.
- Case 6: Pituitary adenoma (secreting). Prescription dose 25Gy. Report maximum dose to the chiasm – $D_{0.030cc}$.
- Case 7: 50 year-old female with HER2 breast cancer. Brain MRI performed following headaches demonstrated 14 small lesions, consistent with metastases. Prescription dose for PTV1-14: 24Gy (GK, CK and Zap-X), 21Gy (C-arm linacs).

These test cases were selected to represent a wide range of intracranial targets, including those adjacent to, or within, OARs. The 28 targets amongst the seven patients ranged from 0.02cc to 7.2cc in volume.

Participating centres were sent DICOMRT files containing images, target contours and potential OARs for each treatment plan. They were asked to plan each treatment to the best of

their ability using experienced staff (defined as at least two years' experience with the relevant platform).

To enable easier comparison of treatment plans, doses were specified for each target. Following common practice, C-arm linacs added a 1mm margin for metastases but no margin for benign disease. Prescription doses were reduced if adding a margin to reduce additional toxicity. Tolerance doses to OARs were agreed in advance. Centres were told not to modify any of the structures, though they were allowed to add additional volumes for optimisation, sparing etc. if they wished.

For metastatic targets, users were asked to achieve ≥99% coverage of the PTV with the prescription dose. For benign targets this was reduced to ≥95% (ideally 97%). Secondary planning objectives included selectivity, gradient and OAR doses.

Parameters used for comparison included coverage, selectivity, Paddick Conformity Index[8], Gradient Index[9], R50%[10], the Efficiency Index and Global Efficiency Index[11], doses to OARs, estimated planning time and estimated treatment time. Calculation formulas can be found in Supplemental material 3.

The Efficiency index is a relatively new metric used to quantify plan quality, including for multiple metastases[11]. It combines conformity, gradient and mean dose within the target into a single value which represents how much of a plan's integral dose is being usefully employed in irradiating the target(s).

Volume-averaged data was calculated for the multiple metastases cases, with the contribution of each individual target weighted according to its volume. This reduced the importance of small mets compared to larger lesions within the same patient.

A Kruskall-Wallis test was performed to indicate significant differences between platforms for the PCI, GI, EI, planning and treatment times. If confirmed, a Mann-Whitney U test was used to compare each platform's results against those of the Gamma Knife. A p-value of <0.05 was chosen to denote significance.

As the study involved the use of historical and anonymised images and contours, approval from an ethics committee was not needed.

Results

Coverage, selectivity, PCI, GI, and R50%, averaged over all lesions, are shown in Table1. As good conformity and fall-off are harder to achieve for small targets, volume averaged data is also provided for the metastasis cases, in order to reduce potential bias from smaller, potentially less clinically critical, targets.

Target coverage

All platforms achieved ≥98% coverage of metastatic lesions and ≥95% for benign targets (SFig1, Supplemental materials 4).

Selectivity

Selectivity is graphically represented in SFig1. Due to bridging of the PIV it was not possible to calculate the selectivity for two pairs of proximal lesions (Case 7, PTVs 4&5, 9&10) so these were removed from the analysis for all platforms.

Paddick Conformity Index

Mean PCIs for benign lesions were highest for EV (0.891) and for metastatic lesions, CK (0.896) (Table 1). The lowest was for Zap-X when planning metastases (0.700). The largest increase in PCI between the average and the volume-weighted average was seen for GK

reflecting the relatively lower conformity for smaller targets planned for the GK compared with larger targets. CK and HA had significantly higher values than GK while Zap plans were significantly lower.

Gradient index

GIs for benign lesions ranged from a mean of 2.74 (Zap-X), representing the steepest dose gradient, to 4.18 (EE). For metastases, GK had the lowest GI at 3.66 and HA-10X the highest at 5.27 ($p<0.001$). Volume-averaging lowered the GI for all platforms but especially for the platforms with relatively high GI values for the smallest mets (CK and HA-10X) (Table 1).

R50%

R50% values (Table 1), which represent a combination of conformity and gradient, had a minimum mean value for benign lesions of 3.12 (Zap-X). The lowest R50% for metastases was GK (4.74) and the highest HA-10X (6.19) ($p=0.027$).

Efficiency Index

For benign targets GK and Zap-X achieved the highest EI values, followed by CK then the C-arm linacs though this didn't reach statistical significance. For metastasis cases, higher Global EIs were seen for GK and CK, with the lowest value scored by Zap-X for Case 1 and HA-10X for Cases 2 and 7 (Fig2).

OAR doses

OAR doses for the benign lesions are presented in SFig3 (Supplemental material 4). For Case 3 all platforms achieved a mean cochlea dose of <4Gy. Most C-arm linacs failed to meet this constraint in the original 2016 benchmark study. Maximum ($D_{0.03cc}$) brainstem doses for Cases 4 and 5 were broadly similar across all platforms. The lowest chiasm maximum dose

($D_{0.03cc}$) for the pituitary plan was achieved by EV (2.6Gy) and the highest was the HA-10X plan (6.5Gy), which may be a consequence of the fixed beam geometry of the HA plans. However, the EE solution had similar chiasm doses along with the lowest conformity, perhaps due to the use of 5mm MLC leaves in combination with the complex shape and proximity of the OAR.

Comparison to the 2016 study

For the three centres participating in both the 2016 and the 2022 benchmarking studies, CK demonstrated a 14% increase in PCI ($p<0.001$) and reduction of 11% in GI ($p=0.070$), while EV showed a 16% increase in PCI ($p<0.001$) and 8% reduction in GI ($p=0.374$). Changes in GK plan quality were more variable with some PCI values decreasing (Fig3).

Planning and delivery time

Fig4 shows the self-reported planning and estimated delivery times for each case across the platforms. Treatment times for CK, EE and EV included intrafraction imaging.

Discussion

In the field of SRS, where the technical capabilities of radiotherapy platforms are pushed to their limit, comparison studies have been very useful. However, bias and poor study design have led to inconsistent outcomes and have led to questioning the utility of such studies[12]. The 2016 study, coordinated by RTTQA consisted of 24 treatment platforms which was more than in any previously published benchmarking study. With all centres motivated to produce the highest quality plans possible, within a realistic timeframe, this created a valuable snapshot of the practice and capabilities of SRS delivery in the UK. Variation in plan quality was surprisingly large but it could be partly explained by the difference in age of the

equipment used as well as wide variations in local practice. This current study selects six contemporary devices and uses experienced treatment planning teams in order to create a 'level playing field'. The following observations can be made regarding the plan quality metrics:

Gradient Index

The GI appears to follow a trend with beam energy, with the lowest values from the lower energy platforms (GK; 1.25MeV, Zap-X; 3MV) and the highest value from the highest energy (HA-10X). This is also seen within the same platform with a HA-6X GI of 4.63 for metastases, compared to 5.27 for 10X. This finding is consistent with other studies[13], and reflects the reduced scatter penumbra for lower energies. Collimation, source-axis-distance (SAD) and prescription isodose level[14] may also play a part in the differences seen in the dose gradient outside the target. The SAD for GK and Zap-X are ~40cm and 45cm respectively, having collimation closer to the patient compared with other platforms, reducing the geometric beam penumbra. Lower prescription isodoses (i.e. greater inhomogeneity within the target) are also typical of GK and Zap-X treatments. CK has an SAD of 80cm compared to 100cm for the C-arm linac-based platforms but also typically delivers treatments to a lower prescription isodose hence an intermediate GI.

PCI v gradient

The trade-off between PCI and GI can be a function of the delivery approach. Elements uses dynamic conformal arcs (DCA) for single isocentre multiple metastases treatments, but VMAT for single lesions. As metastases are typically spherical and usually distant from OARs, DCA allow fast delivery, simple dosimetry and lower MU. DCA offer less opportunity to conform a high dose around the target but may deliver a higher dose fall-off than a VMAT solution[15,16]. This was reflected in EV being the only C-arm linac plans with

GIs that were not significantly higher than GK. The balance between PCI and GI for all lesions is illustrated in Figure 2d, where it can be seen that the VMAT based solution of HA has higher conformity than the DCA based BrainLab Elements solution but at the expense of a higher GI.

Both Elements and Hyperarc centres used VMAT for benign targets, and in general there is less difference between platforms apart from Case 3, a 0.06cc acoustic target where the 5mm MLC leaf width may reduce plan quality for the EE solution.

Efficiency Index

The EI is generally highest for GK and Zap-X, with CK treatments also yielding average indices that are higher than the C-arm linacs. This is likely due to a combination of an average higher internal dose and steeper gradient (GK and Zap-X) or high selectivity (CK). Interestingly, while Zap-X scores overall highest for benign targets, it has the lowest value of any device for Case 1. This is possibly due to decisions made when treatment planning, balancing conformity, dose fall-off and overall treatment time.

Planning time

Reported planning times range from 10 minutes to plan Case 4 for GK to 4 hours to plan Case 7 for CK. All centres spend longer on average planning the multiple metastasis cases than the benign cases. The shortest total planning time is 120 minutes for GK, and the longest 960 minutes for CK ($p=0.001$). Planning times for the highly automated C-arm linac-based solutions are broadly similar.

Planning time is self-reported, and reflects some elements of user preference. Even with the same planning system, differences in approach between centres can be seen, with the EE centre spending less time per plan than the EV centre. This may be partly due to the EE

centre using a newer version of the TPS with increased automation, and faster computational speeds, but may also reflect how much time individual planners spend exploring options to improve plan quality once an acceptable plan has been achieved.

Treatment time

There is a notable difference in the total treatment times between the C-arm linac platforms and the specialist platforms. Differences are most marked for the multiple brain mets cases (1,2,7) where the ability to treat multiple mets with a single isocentre on the linac-based platforms reduces delivery times compared to the multiple isocentre delivery methods of GK/CK/Zap-X. The biggest difference in any individual case was seen for the 14-met plan (Case 7), which took 204m to deliver on GK compared to 12.1m using HA-10X.

The shortest total delivery time for the 7 cases for a C-arm linac-based platform is 68m for HA-10X and the longest 210m for EV, with differences mainly due to the use of 3 separate isocentres for the 14-met plan and the choice of a flattened beam with lower dose rate (2400MU/min vs 600MU/min). However, there was minimal difference in treatment times for EE versus EV despite EE delivering treatments with an FFF beam. Timings for all linac-based solutions included an initial CBCT scan but the Elements timings also included intrafraction imaging at each couch angle using ExacTrac.

Delivery timings were much higher for the specialist platforms, ranging from 579m for CK to 695m for GK. The latter depends on the age of the Co-60 sources and these were normalised in this study to a reference dose rate of 3.0Gy/min. For mask-based treatments, treatment times were calculated assuming that treatment wasn't paused to account for patient movement.

GK, CK and Zap-X used multiple isocentres for the multiple met cases (Case 1,2,7). Three of the linac-based platforms used a single isocentre set-up for the 3-met (Case 1) and 7-met (Case 2) plans, whilst EE chose to treat the Case 1 brainstem met with a separate isocentre. This is their preferred clinical solution to minimise any set-up error in the brainstem met (as set-up errors increase with distance off-axis for single isocentre multiple met plans)[17]. Both Elements platforms used 3 isocentres to treat the 14-met case. Elements can create a single isocentre plan treating all 14 mets but clinical practice at both centres appears to align, in that clusters of mets which are at a distance from the others and/or in close proximity to each other are treated separately in order to minimise target distance from the isocentre, and give more opportunity to minimise potential bridging doses between pairs of very close targets. Treating with additional isocentres can also minimise the total dose to normal brain. In the case of the 14 met plan, total normal brain V12Gy for the combined 3 isocentres is 19.6cc for both Elements centres, compared to 25.8cc (HA-6X) and 26.3cc (HA-10X) although at the expense of increased treatment time.

Treatment times for the complex shaped benign lesions were also longer on the specialist platforms, in comparison to the C-arm linacs. These times are not meant to be targets or limits, rather the aim is to show the general balance per platform between planning and delivery times.

Plan improvements

Improvements in plan quality are illustrated in Figure 3 by comparing the individual plan metrics for the 3 centres who participated in the 2016 benchmarking exercise[1]. Both CK and EV plans show significant improvements, which may be owed to their new respective planning systems. Changes are more variable for the GK centre and in some cases the new GK plans have notably inferior conformity. This may be explained by the use of the

Lightning inverse planning algorithm which, while extremely fast, can, in some circumstances, produce plans that are inferior to an expert manual treatment planner[18]. However, it should be noted that this platform had the highest original plan quality. In general, differences in plan quality between the platforms appear to have narrowed since the original benchmarking exercise in 2016.

Limitations

Planning metastatic targets differently (in terms of prescription dose and margins applied) between the dedicated vs C-arm linac platforms, while creating difficulty in a direct performance comparison across all devices, allowed a closer comparison between the actual treatments typically delivered with these devices.

The addition of a margin to a target effectively smooths its contours and also increases its volume. For small metastases this volume increase can be several-fold and can add bias to the comparison of margin vs no-margin plans, particularly for those indices that tend to vary with volume. For benign targets, where no margin and an identical prescription dose has been used for all platforms, variations in coverage, selectivity and conformity are narrower, while gradient still appears to be dependent on beam energy, with Co-60 and 3MV energies and shorter SAD/closer collimation having a steeper gradient than 6 and 10X. R50% values also demonstrate a greater variation amongst platforms.

Variation in the prescription isodose, which governs the maximum dose inside the target, but also affects the gradient outside the target, has not been evaluated. Clinical studies have shown little evidence that the internal maximum dose has a clinical effect[19] while the gradient, which was analysed, is known to affect the risk of symptomatic radionecrosis[20–22].

In general, local TPS values were used, which may lead to bias due to the different methods of target volume and dose-volume calculation. These uncertainties may amount to 5-10%[23,24], and should be borne in mind when appraising small differences between platforms, but are unlikely to change the overall trends and conclusions in this study. The spacing of dose calculation points also vary between planning systems with GK using 0.5mm and Zap-X, CK and C-arm linacs using 1.0mm. Typically, a finer dose calculation grid will lead to slightly poorer quality parameters being calculated. In addition, the Zap TPS calculates volumes to the nearest 0.01cc while all others calculate to the nearest 0.001cc. This can introduce rounding errors of up to 1% when assessing the PCI of a lesion of 1.0cc and up to a 10% error for targets of 0.1cc. PTV volumes as small as 0.02cc are present in this study so TPS calculated indices of the smaller targets should be viewed with caution.

Due to the low number of benign targets planned, it was not possible to reach statistical significance in any parameter for this group. In addition, the large number of relationships that could have been explored required limitation so we opted to compare all platforms to the Gamma Knife which is often seen as the 'gold standard' in SRS.

The Zap-X system is still in its infancy and so finding a centre with more than two years' experience was not possible. A staff member of Zap Surgical was consulted to create plans that were as close as possible to that expected from a clinical centre.

We present data from a limited number of centres from one country. The results obtained may not be reflective of the case mix or clinical practice in other regions. However, our study used a range of metastatic and benign cases, whereas some planning studies have relied on a single case[25], or malignant or benign cases only[26,27].

All plan comparison studies present an incomplete picture of treatment quality, however they can be useful to inform current and future practice[12]. This study has not considered delivery

accuracy or the clinical significance of the differences seen. However, it is intended to give a picture of contemporary clinical practice using the latest treatment platforms, assessed using comparable and consistent parameters.

Conclusion

The national 2016 SRS benchmark study has been repeated and expanded upon, with representative centres for each of the main contemporary SRS platforms.

Our results suggest a notable improvement in plan quality for CK and C-arm linacs as a result of technological improvements of both treatment planning systems and hardware. This has led to a reduced variability between different platforms compared with that seen in the original 2016 benchmarking study.

Figure 1. a) PCI versus target volume, b) GI versus target volume; c) R50% versus target volume; d) GI versus PCI. Results for the smallest volumes calculated for charts a-c are subject to large uncertainties, hence the absence of multiple horizontal points on the x-axis. A 1mm margin can increase the volume by up to an order of magnitude, shifting C-arm linac points to the right.

Figure 2. (left) Global EI for the multiple metastasis cases. (right) EI for the benign cases

Figure 3. PCI v GI for Cases 1 and 2 for three centres (Gamma Knife, CyberKnife and Varian/Brainlab). Open markers show the 2016 data points, solid markers the 2022 data and the arrows the direction and magnitude of change.

Figure 4. Self-reported planning (orange) and estimated treatment times (blue) for each case.

References

1. Eaton DJ, Lee J, Paddick I. Stereotactic radiosurgery for multiple brain metastases: Results of multicenter benchmark planning studies. *Pract Radiat Oncol*. 2018;8(4):e212-e220. https://doi:10.1016/j.prro.2017.12.011

2. Eaton DJ, Lee J, Patel R, Millin AE, Paddick I, Walker C. Stereotactic radiosurgery for benign brain tumors: Results of multicenter benchmark planning studies. *Pract Radiat Oncol*. 2018;8(5):e295-e304. https://doi:10.1016/j.prro.2018.02.006

3. Lindquist, Christer M, Paddick, Ian Ms. The Leksell Gamma Knife Perfexion and Comparisons with Its Predecessors. *Neurosurgery*. 2007;61(September):130-141. https://doi:10.1227/01.NEU.0000279989.86544.B6

4. Sjölund J, Riad S, Hennix M, Nordström H. A linear programming approach to inverse planning in Gamma Knife radiosurgery. *Med Phys*. 2019;46(4):1533-1544. https://doi:10.1002/mp.13440

5. Adler JRJ, Chang SD, Murphy MJ, Doty J, Geis P, Hancock SL. The Cyberknife: A Frameless Robotic System for Radiosurgery. *Stereotact Funct Neurosurg*. 1997;69:124-128. https://doi:10.1159/000099863

6. Kilby W, Naylor M, Dooley J, Maurer JC, Sayeh S. *A Technical Overview of the CyberKnife System. Handbook of Robotic and Image-Guided Surgery.* (Abedin-Nasab M, ed.). Elsevier; 2020.

7. Srivastava SP, Jani SS, Pinnaduwage DS, et al. Treatment planning system and beam data validation for the ZAP-X: A novel self-shielded stereotactic radiosurgery system. *Med Phys*. 2021;48(5):2494-2510. https://doi:10.1002/mp.14740

8. Paddick I. A simple scoring ratio to index the conformity of radiosurgical treatment plans. Technical note. *J Neurosurg*. 2000;93(SUPPL. 3):219-222. https://doi:10.3171/jns.2000.93.supplement_3.0219

9. Paddick I, Lippitz B. A simple dose gradient measurement tool to complement the conformity index. *J Neurosurg*. 2006;105 Suppl:194-201. https://doi:10.3171/sup.2006.105.7.194

10. Timmerman R, Paulus R, Galvin J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. 2010;303(11):1070-1076. https://doi:10.1001/jama.2010.261.Stereotactic

11. Dimitriadis A, Paddick I. A novel index for assessing treatment plan quality in stereotactic radiosurgery. *J Neurosurg*. 2018;129(December):118-124. https://doi:10.3171/2018.7.GKS18694

12. Eaton DJ. Why all radiotherapy planning studies are wrong but some are useful. *Medical Dosimetry*. 2021;46(3):209-211. https://doi:10.1016/j.meddos.2020.12.002

13. Laoui S, Roa D E, Kuo J, et al. Flattening filter free beam energy selection and its impact in multitarget intracranial stereotactic radiosurgery treatments. *Med Dosim*. 2020;45(4):363-367. Accessed May 14, 2021. https://doi.org/10.1016/j.meddos.2020.05.001

14. Zhao B, Jin JY, Wen N, et al. Prescription to 50-75% isodose line may be optimum for linear accelerator based radiosurgery of cranial lesions. *J Radiosurg SBRT*. 2014;3(2):139-147.

15. Huang Y, Chin K, Robbins JR, et al. Radiosurgery of multiple brain metastases with single-isocenter dynamic conformal arcs (SIDCA). *Radiotherapy and Oncology*. 2014;112(1):128-132. https://doi:10.1016/j.radonc.2014.05.009

16. Skourou C, Hickey D, Rock L, et al. Treatment of multiple intracranial metastases in radiation oncology: a contemporary review of available technologies. *BJR|Open*. 2021;3(1):20210035. https://doi:10.1259/bjro.20210035

17. Roper J, Chanyavanich V, Betzel G, Switchenko J, Dhabaan A. Single-Isocenter Multiple-Target Stereotactic Radiosurgery: Risk of Compromised Coverage. *Int J Radiat Oncol Biol Phys*. 2015;93(3):540-546. https://doi:10.1016/j.ijrobp.2015.07.2262

18. Paddick I, Grishchuk D, Karanatsiou A. Automatic planning (Lightning) evaluation for Gamma Knife radiosurgery for AVMs: Comparison with Manual planning and IntuitivePlan inverse planning. *J Radiosurg SBRT*. 8:116-117.

19. Sneed P K, Mendez J, Vemer-van den Hoek J G M, et al. Adverse radiation effect after stereotactic radiosurgery for brain metastases: incidence, time course, and risk factors. *J Neurosurg*. 2015;123:373-386. Accessed May 14, 2021. https://doi.org/10.3171/2014.10.JNS141610

20. Flickinger JC, Kondziolka D, Dade Lunsford † L, et al. Development of a model to predict permanent symptomatic postradiosurgery injury for arteriovenous malformation patients. *Int J Radiat Oncol Biol Phys*. 2000;46(5):1143-1148. https://doi.org/10.1016/S0360-3016(99)00513-1

21. Korytko T, Radivoyevitch T, Colussi V, et al. 12 Gy gamma knife radiosurgical volume is a predictor for radiation necrosis in non-AVM intracranial tumors. *Int J Radiat Oncol Biol Phys*. 2006;64(2):419-424. https://doi:10.1016/j.ijrobp.2005.07.980

22. Minniti G, Clarke E, Lanzetta G, et al. Stereotactic radiosurgery for brain metastases: Analysis of outcome and risk of brain radionecrosis. *Radiation Oncology*. 2011;6(1):1-9. https://doi:10.1186/1748-717X-6-48

23. Eaton DJ, Alty K. Dependence of volume calculation and margin growth accuracy on treatment planning systems for stereotactic radiosurgery. *British Journal of Radiology*. 2017;90(1080). doi:10.1259/bjr.20170633

24. Ma L, Sahgal A, Nie K, et al. Reliability of contour-based volume calculation for radiosurgery. *J Neurosurg*. 2012;117 Suppl(December):203-210. doi:10.3171/2012.7.gks121016

25. Chan MKH, Gevaert T, Kadoya N, et al. Multi-center planning study of radiosurgery for intracranial metastases through Automation (MC-PRIMA) by crowdsourcing prior web-based plan challenge study. *Physica Medica*. 2022;95(October 2021):73-82. doi:10.1016/j.ejmp.2022.01.011

26. Chea M, Fezzani K, Jacob J, et al. Dosimetric study between a single isocenter dynamic conformal arc therapy technique and Gamma Knife radiosurgery for multiple brain metastases treatment: impact of target volume geometrical characteristics. *Radiation Oncology*. 2021;16(1):1-16. https://doi:10.1186/s13014-021-01766-w

27. Gevaert T, Levivier M, Lacornerie T, et al. Dosimetric comparison of different treatment modalities for stereotactic radiosurgery of arteriovenous malformations and

acoustic neuromas. *Radiother Oncol*. 2013;106:192-197. Accessed May 13, 2021. http://www.ncbi.nlm.nih.gov/pubmed/22884842

Figure 1

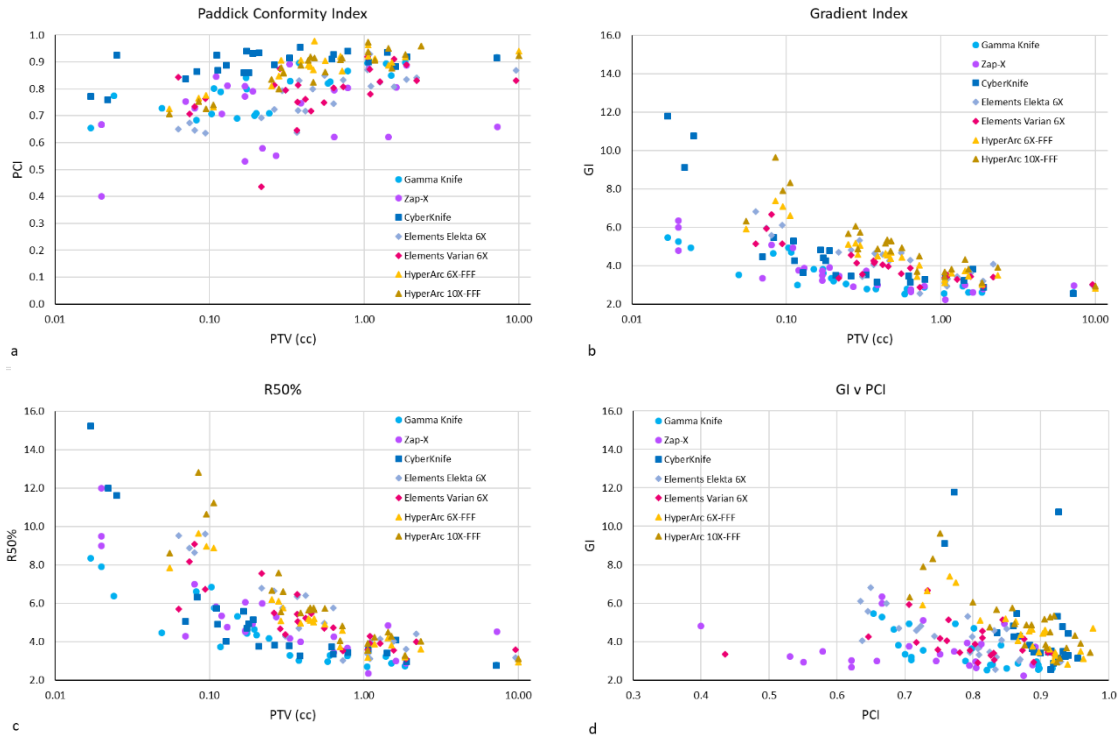
Paddick Conformity Index
PCI
PTV (cc)
Gamma Knife
Zap-X
CyberKnife
Elements Elekta 6X
Elements Varian 6X
HyperArc 6X-FFF
HyperArc 10X-FFF
a
Gradient Index
GI
PTV (cc)
b
R50%
R50%
PTV (cc)
c
GI v PCI
GI
PCI
d

Figure 2

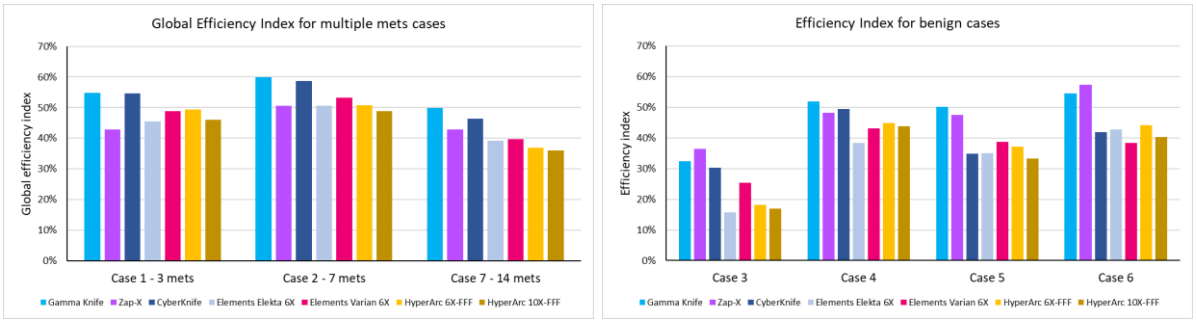
Global Efficiency Index for multiple mets cases
Global efficiency index
70%
60%
50%
40%
30%
20%
10%
0%
Case 1 - 3 mets
Case 2 - 7 mets
Case 7 - 14 mets
Gamma Knife
Zap-X
CyberKnife
Elements Elekta 6X
Elements Varian 6X
HyperArc 6X-FFF
HyperArc 10X-FFF
Efficiency Index for benign cases
Efficiency index
70%
60%
50%
40%
30%
20%
10%
0%
Case 3
Case 4
Case 5
Case 6
Gamma Knife
Zap-X
CyberKnife
Elements Elekta 6X
Elements Varian 6X
HyperArc 6X-FFF
HyperArc 10X-FFF

Figure 3

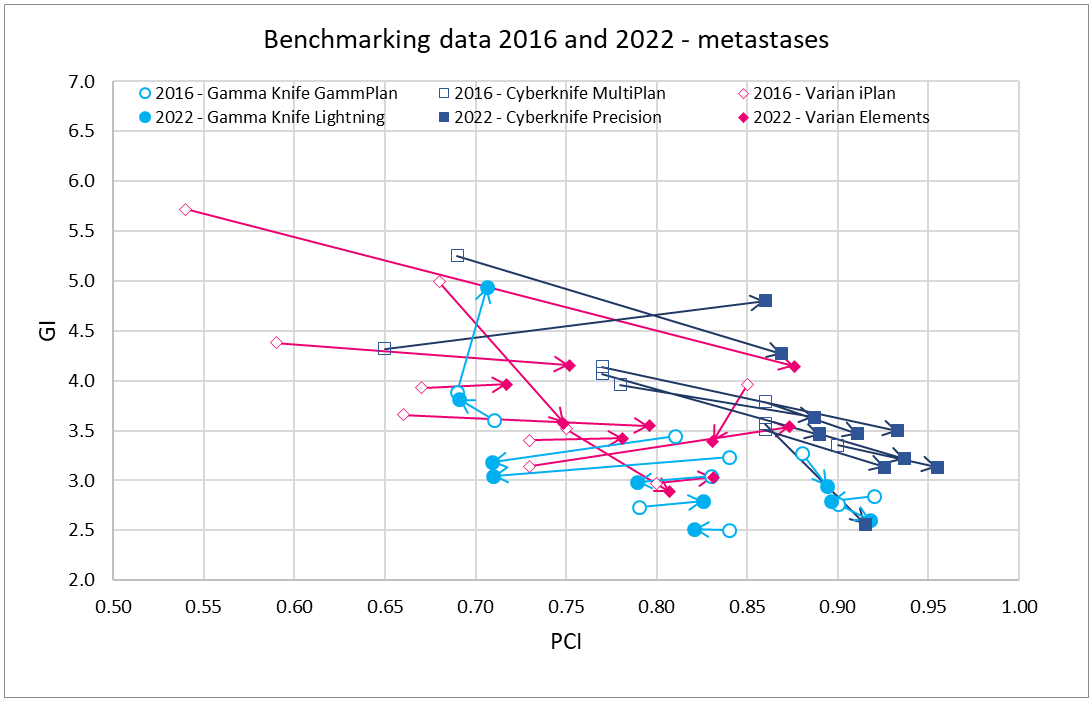
Benchmarking data 2016 and 2022 - metastases
2016 - Gamma Knife GammPlan
2016 - Cyberknife MultiPlan
2016 - Varian iPlan
2022 - Gamma Knife Lightning
2022 - Cyberknife Precision
2022 - Varian Elements
GI
PCI
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00

Figure 4

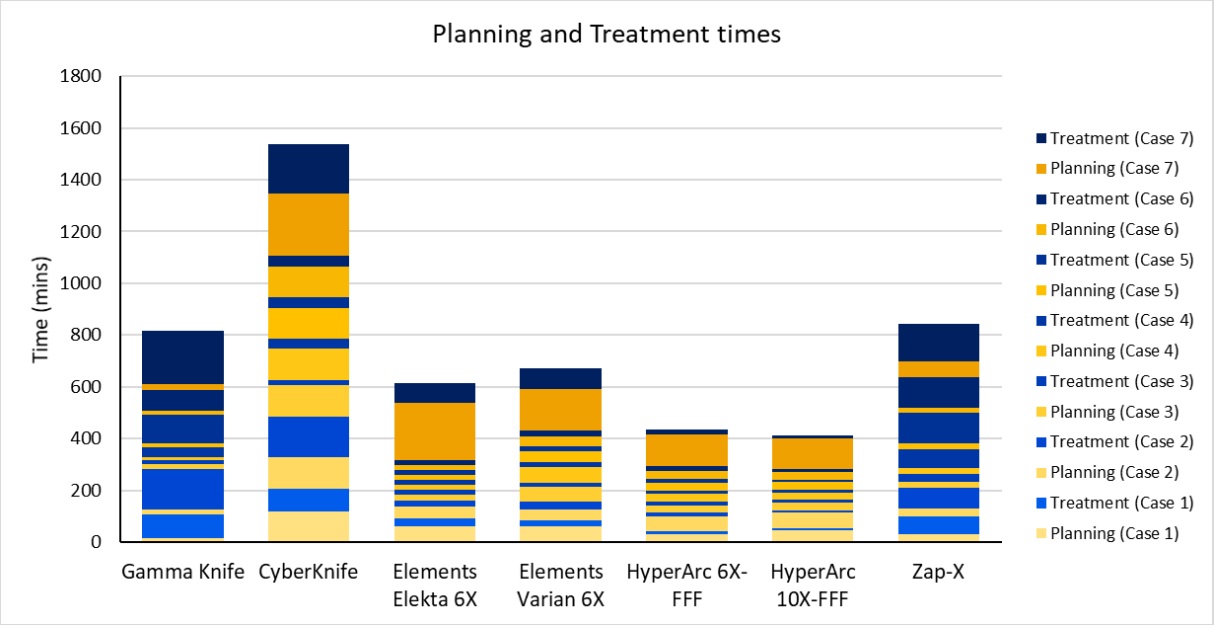
Planning and Treatment times
Time (mins)
1800
1600
1400
1200
1000
800
600
400
200
0
Gamma Knife
CyberKnife
Elements Elekta 6X
Elements Varian 6X
HyperArc 6X-FFF
HyperArc 10X-FFF
Zap-X
Treatment (Case 7)
Planning (Case 7)
Treatment (Case 6)
Planning (Case 6)
Treatment (Case 5)
Planning (Case 5)
Treatment (Case 4)
Planning (Case 4)
Treatment (Case 3)
Planning (Case 3)
Treatment (Case 2)
Planning (Case 2)
Treatment (Case 1)
Planning (Case 1)

Table 1. Coverage, Selectivity, PCI, GI and R50% for each platform, averaged for benign and metastases separately. For the metastasis cases, volume averaged data is shown in brackets. Statistical significance relative to values from the Gamma Knife, are shown in italics. P-values <0.05 are shown in bold.

*Mets 4,5,9 & 10 of Case 7 were excluded from the analysis.

	Gamma Knife	Zap-X	CyberKnife	Elements Elekta 6X-FFF	Elements Varian 6X	HyperArc 6X-FFF	HyperArc 10X-FFF
	Benign lesions						
Coverage	96.7%	96.2%	97.1%	96.1%	97.5%	98.9%	99.1%
Selectivity	0.871	0.862	0.911	0.822	0.914	0.886	0.878
PCI	0.843	0.830	0.885	0.790	0.891	0.876	0.871
GI	2.83	2.74	3.64	4.18	3.72	3.87	4.17
R50%	3.19	3.12	3.91	5.13	3.99	4.50	4.93
	Metastases* (volume averaged)						
Coverage	99.0% (98.8)	100.0% (99.9)	99.7% (99.4)	98.2% (98.3)	99.1% (99.0)	99.8% (99.7)	99.6% (99.6)
Selectivity	0.787 (0.862)	0.700 (0.745)	0.899 (0.928)	0.784 (0.841)	0.773 (0.818)	0.878 (0.913)	0.873 (0.919)
PCI	0.779 (0.852)	0.700 (0.744)	0.896 (0.922)	0.768 (0.825)	0.766 (0.809)	0.877 (0.911)	0.870 (0.915)
p-value	-	***0.019***	***<0.001***	*0.420*	*0.490*	***<0.001***	***<0.001***
GI	3.66 (2.91)	3.83 (3.18)	4.89 (3.37)	4.31 (3.49)	4.05 (3.48)	4.63 (3.71)	5.27 (4.11)
p-value	-	*0.249*	***0.036***	***0.030***	*0.057*	***0.009***	***<0.001***
R50%	4.74 (3.38)	5.69 (4.31)	5.57 (3.63)	5.60 (4.21)	5.32 (4.28)	5.37 (4.10)	6.19 (4.50)
p-value	-	***0.040***	*0.299*	*0.086*	*0.074*	*0.104*	***0.027***

Abbreviations: FFF = Flattening Filter-Free, PCI = Paddick Conformity Index, GI = Gradient Index