

2-5-2022

## Multi-center planning study of radiosurgery for intracranial metastases through Automation (MC-PRIMA) by crowdsourcing prior web-based plan challenge study

M K H Chan

T Gevaert

N Kadoya

J Dorr

R Leung

*See next page for additional authors*

Follow this and additional works at: <https://jdc.jefferson.edu/radoncfp>



Part of the [Oncology Commons](#), and the [Radiation Medicine Commons](#)

**[Let us know how access to this document benefits you](#)**

---

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Radiation Oncology Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).

---

**Authors**

M K H Chan, T Gevaert, N Kadoya, J Dorr, R Leung, S Alheet, A Toutaoui, R Farias, M Wong, C Skourou, M Valenti, I Farré, C Otero-Martínez, D O'Doherty, J Waldron, S Hanvey, M Grohmann, and Haisong Lui

---



## Original paper

## Multi-center planning study of radiosurgery for intracranial metastases through Automation (MC-PRIMA) by crowdsourcing prior web-based plan challenge study

M.K.H. Chan<sup>a,\*</sup>, T. Gevaert<sup>b</sup>, N. Kadoya<sup>c</sup>, J. Dorr<sup>d</sup>, R. Leung<sup>e</sup>, S. Alheet<sup>f</sup>, A. Toutaoui<sup>g</sup>, R. Farias<sup>h</sup>, M. Wong<sup>i,j</sup>, C. Skourou<sup>k</sup>, M. Valenti<sup>l</sup>, I. Farré<sup>m</sup>, C. Otero-Martínez<sup>n</sup>, D. O'Doherty<sup>o</sup>, J. Waldron<sup>p</sup>, S. Hanvey<sup>q</sup>, M. Grohmann<sup>r</sup>, H. Liu<sup>s</sup>

<sup>a</sup> Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, the Netherlands

<sup>b</sup> Department of Radiotherapy, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Belgium

<sup>c</sup> Department of Radiation Oncology, Tohoku University Graduate School of Medicine, Japan

<sup>d</sup> AdventHealth, USA

<sup>e</sup> Auckland Radiation Oncology, New Zealand

<sup>f</sup> Department of Radiation Oncology, King Hussein Cancer Center, Jordan

<sup>g</sup> Chahids Mahmoudi Hospital, Algeria

<sup>h</sup> Department of Radiation Oncology, Mevaterapia Medical Center, Argentina

<sup>i</sup> GenesisCare, Cabrini Hospital, Australia

<sup>j</sup> Department of Clinical Oncology, Tuen Mun Hospital, China

<sup>k</sup> Department of Physics, St. Luke's Radiation Oncology Network, Ireland

<sup>l</sup> Unit of Medical Physics-Ospedali Riuniti Umberto I, G.M. Lancisi, G. Salesi Ancona, Italy

<sup>m</sup> Servei de Física Mèdica i Protecció Radiològica, Institut Català d'Oncologia, Spain

<sup>n</sup> Department of Medical Physics and Radiation Protection, Hospital Clínico Universitario of Santiago de Compostela, Spain

<sup>o</sup> Cambridge University Hospitals NHS Foundation Trust, UK

<sup>p</sup> Newcastle Upon Tyne Hospitals NHS Foundation Trust, UK

<sup>q</sup> University Hospitals Plymouth NHS Trust, UK

<sup>r</sup> Department of Radiotherapy and Radiation Oncology, University Medical Center Hamburg-Eppendorf, Germany

<sup>s</sup> Department of Radiation Oncology, Thomas Jefferson University, USA



## ARTICLE INFO

## Keywords:

Plan crowdsourcing  
Autoplanning  
Stereotactic radiosurgery  
Multiple brain metastases

## ABSTRACT

**Background:** Planning radiosurgery to multiple intracranial metastases is complex and shows large variability in dosimetric quality among planners and treatment planning systems (TPS). This project aimed to determine whether autoplanning using the Multiple Brain Mets (AutoMBM) software can improve plan quality and reduce inter-planner variability by crowdsourcing results from prior international planning study.

**Methods:** Twenty-four institutions autoplanned with AutoMBM on a five metastases case from a prior international planning competition from which population statistics (means and variances) of 23 dosimetric metrics and resulting composite plan score (maximum score = 150) of other TPS (Eclipse, Monaco, RayStation, iPlan, GammaPlan, MultiPlan) were crowdsourced. Plan results of AutoMBM and each of the other TPS were compared using two sample *t*-tests for means and Levene's tests for variances. Plan quality of AutoMBM was correlated with the planner's experience and compared between academic and non-academic centers.

**Results:** AutoMBM produced plans with comparable composite plan score to GammaPlan, MultiPlan, Eclipse and iPlan (127.6 vs. 131.7 vs. 127.3 vs. 127.3 and 126.7; all  $p > 0.05$ ) and superior to Monaco and RayStation (118.3 and 108.6; both  $p < 0.05$ ). Inter-planner variability of overall plan quality was lowest for AutoMBM among all TPS (all  $p < 0.05$ ). AutoMBM's plan quality did not differ between academic and non-academic centers and uncorrelated with planning experience (all  $p > 0.05$ ).

**Abbreviations:** AP, autoplanning; TPS, treatment planning system; SRS, stereotactic radiosurgery; DCA, dynamic conform arc; VMAT, volumetric modulated arc radiotherapy; MLC, multi-leaf collimator; PCI, Paddict conformity index; GI, dose gradient index; R50%, spread of half isodose line;  $V_x$  Gy, volume receiving  $x$  Gy or more;  $D \times \text{cm}^3$ , dose to  $x \text{ cm}^3$  of the volume.

\* Corresponding author at: Department of Radiation Oncology, University Medical Center Groningen, P.O. Box 30001, 9700 Groningen, the Netherlands.

E-mail address: [k.h.chan@umcg.nl](mailto:k.h.chan@umcg.nl) (M.K.H. Chan).

<https://doi.org/10.1016/j.ejmp.2022.01.011>

Received 11 October 2021; Received in revised form 10 January 2022; Accepted 28 January 2022

Available online 5 February 2022

1120-1797/© 2022 Associazione Italiana di Fisica Medica e Sanitaria. Published by Elsevier Ltd. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

*Conclusions:* By plan crowdsourcing prior international plan challenge, AutoMBM produces high and consistent plan quality independent of the planning experience and the institution that is crucial to addressing the technical bottleneck of SRS to intracranial metastases.

## Introduction

Historically, stereotactic radiosurgery (SRS) was offered only to treat limited number of lesions [1]. Recent evidences have been established to show safety of SRS for patients with more than four lesions and even beyond ten without compromised overall survival and increasing incidence of SRS-related adverse events [2–5].

Besides the prohibitively long treatment duration associated with treating a large number of lesions, each of which may require additional shots for GammaKnife and CyberKnife, or in the case of C-arm linac a separate isocenter, the other major challenge is concerned with treatment planning. Planners typically rely on experience to determine and iteratively refine a range of treatment platform-and delivery technique-specific variables. A particular aspect of this exhaustive process is to minimize the dose bridging between lesions, which leads to unwanted high dose in the surrounding normal tissues. As the number of lesions increases, the complexity of finding an optimal set of variables by, for example, adjusting the size and shape of the collimation aperture and the irradiation trajectory (e.g., beam incident angles) can rapidly expand beyond the capacity of human planner. The planner's experience eventually becomes the determinant of the plan quality and main driver of plan variability, which might not be adequately compensated with the aid of general inverse-optimizers.

On the other hand, standardization of plan quality might not necessarily be achieved through plan benchmarking, as shown by the UK National Health Service commissioning of SRS service [6]. Despite provision of detailed protocol, guidelines and feedbacks, for two plan benchmark cases comprising three and seven lesions marked variations of the SRS plans from 21 institutions were still noted not just between treatment platforms but also within the same type of treatment platform. The variation of plan quality was found to be particularly large among institutions that deployed C-arm linacs for SRS indicative of strong dependence of planner's skill.

Automated planning (AP) through scripts, templates or plan library solutions has demonstrated its potential to improve plans throughout and plan quality in many disease sites [7–9]. Among a few commercially available AP solutions, only two were developed with dedicated algorithms to address the unique problem of plan optimization for multiple brain metastases. In a dual-center study, Haisong et al. found favourable target dose conformity and intermediate dose volume 10 and 8 Gy received by the normal brain in non-AP plans generated on the Eclipse TPS (Varian, Palo Alto, CA, USA) for volumetric modulated arc radiotherapy (VMAT) delivery but worse low dose volume < 5 Gy compared to those autoplanned by the Multiple Brain Mets SRS<sup>TM</sup> (AutoMBM; Brainlab, Munich, Germany) software for dynamic conformal arc (DCA) delivery [10]. In a more recent study by Hofmaier et al., the same AutoMBM as used in Haisong et al. [10] was found to generate plans that outperformed non-AP plans using the Monaco (Elekta, Stockholm, Sweden) TPS for VMAT delivery [11]. Earlier, Gevaert et al. indicated that AutoMBM was able to produce superior target dose conformity and normal brain dose than manually optimized VMAT plans on Eclipse [12]. Vergalaso et al. in another multi-center study found that the DCA plans from AutoMBM produced generally worse target dose conformity than the other AP solution of VMAT by HyperArc<sup>TM</sup> (Eclipse, Varian) and non-AP solution of GammaKnife plans [13]. Nonetheless, all these studies and others [6,14] merely reflect the experiences of academically oriented practices and might not be generalizable to clinical environments where quality control and improvement program are less developed. The level of expertise of the participating institutions, case selection and the limited number of plans that could be

generated in the experiment and the comparison arms, etc., prohibit unbiased results and objective conclusions in the above single and multi-center planning studies.

To gain better insights into the potential of AP for multiple brain metastases, a large scale multi-center study of the full range of planning solutions that reflect a variety of academic and non-academic practices should be performed. However, this requires extensive work to create benchmark plans on multiple TPS with and without AP solutions. Moghanaki et al. have recently adopted a crowdsourcing approach to analyze the quality and variability of a large number of stereotactic body radiotherapy plans for a single case of lung cancer using a web-based platform [15]. Hardcastle et al. investigated the challenges of spinal radiosurgery by crowdsourcing 149 plans of spinal radiosurgery from an international plan challenge study that was organized by the Trans-Tasmania Radiation Oncology Group (TROG) [16]. The same group also crowdsourced 160 plans for a case of SRS to five brain metastases from the other international planning study [17]. None of the 160 submissions to this planning competition was known to be auto-planned by AutoMBM. The crowdsourcing approach offers the possibility to efficiently benchmark plan solution belonging to a certain category of treatment platform, delivery technique or TPS by crowdsourcing the results that existed for the other categories from the cloud-database. The feasibility of adapting the crowdsourcing approach has been demonstrated by Giles et al. to benchmark their in-house developed conformal arc informed VMAT (CAVMAT) solution [18].

This Multi-Center Planning Radiosurgery for Intracranial Metastases through Automation (MC-PRIMA) study adapted the crowdsourcing approach to benchmark the performance of AutoMBM against the same 160 plans in the TROG planning study that were analyzed by Hardcastle et al. [17]. But unlike their study that focused on the plan variability by delivery technique (e.g., GammaKnife, CyberKnife, VMAT, intensity modulated therapy), the primary objective of MC-PRIMA was to assess the potential of AutoMBM to improve the quality and reduce the variability of plans versus other existing TPS without dedicated AP solution to SRS of multiple brain metastases. As secondary objective to understand how AutoMBM may address the technical demands of SRS planning, the correlation of plan quality from AutoMBM with the planning experience and the plan quality between academic and non-academic centers were further evaluated.

## Methods and materials

### Recruitment of participants

This study involved as broad a spectrum of participants as possible to emulate the constitution of participants in the TROG planning competition. The final recruitment of participants included nine non-academic and fourteen academic from seven regions (North / South America  $n = 2/2$ ; Europe  $n = 15$ ; Asia  $n = 2$ ; Middle East  $n = 1$ ; Africa  $n = 1$ ; Australasian  $n = 1$ ). The eligibility of institutions in MC-PRIMA was that either the institution clinically use the autoplanning Multiple Brain Mets SRS<sup>TM</sup> (AutoMBM) software to perform autoplanning (AP) or had undergone training from the vendor to use AutoMBM. Furthermore, these institutions must have prior experiences with other SRS planning solutions. The vendor of AutoMBM was also invited to participate.

### Plan study dataset, planning protocols and plan quality metrics

This study adapted the international planning study case of five brain metastases that was originated from the Trans-Tasman Radiation

Oncology Group (TROG) Local HER 0 trial [19]. This study case was published through the publicly accessible web-based plan challenge study platform ProKnow<sup>1</sup> (Elekta, Stockholm, Sweden). The dataset includes anonymized CT images from the skull apex to the second cervical vertebra and a set of tumor and normal organ contours defined according to the Local HER 0 trial [19]. The resolution of the planning CT is  $1.0 \times 0.468 \times 0.468 \text{ mm}^3$ . These gross tumor volumes (GTV) of sizes 0.52 (GTV1), 0.39 (GTV2), 0.07 (GTV3), 2.82 (GTV4) and 0.12 (GTV5)  $\text{cm}^3$  were defined in the frontal, parietal and temporal lobes and the cerebellum adjacent to the brainstem. The sphericity of the GTV, calculated using the OpenCAD extension to 3DSlicer [20] as a measure of the roundness or spherical nature of the target, is 0.54, 0.59, 0.64, 0.59 and 0.66 for GTV1-5, respectively. The sphericity of a sphere is the maximum value of 1. The smaller the value, the less the target approaches to be spherical.

Other normal organs included in the study case were both eyes, lens, optic nerves, hippocampuses, brainstem, optic chiasm and normal brain (i.e., brain minus all GTVs). Per the Local HER 0 trial, GTVs were treated without safety margin for microscopic disease (i.e., no clinical target volume) and geometric uncertainty (i.e., no planning target volume), and no planning risk volume (PRV) for normal organ was defined.

#### User process

Participants were provided with an instruction which included the planning protocol and a web-link to download the plan dataset. Prior to the start of planning, participants were advised to gain full understanding of the planning protocol and the plan quality metrics. Participants then performed AP using AutoMBM and uploaded the resulting plan to the principal investigators for data analysis. Besides collection of the SRS plans, experiences of the planners were documented.

#### Treatment planning with AutoMBM

The AutoMBM is an automated planning solution to treat multiple brain metastases by mono-isocentric arc delivery with multileaf-collimator (MLC) on C-arm linac. Different from the previous studies that all used the early AutoMBM version 1.5, over half of our participants planned with the later AutoMBM version 2.0 using different C-arm linacs. Common to both versions is the template-driven automated planning process. The planner define a set of templates called Clinical Protocols and Setup Protocols each catering to specific treatment objectives and irradiation geometry, respectively. Since version 2.0, AutoMBM employs an inverse algorithm which actively optimizes target heterogeneity, normal brain dose as well as OAR dose by optimizing collimator angle, field shapes, arc weight (monitor unit) and BEV margins. Besides the change of the optimization algorithm, AutoMBM v.2.0 adapts the highly automated planning approach to enable the users in graphical interface to explicitly adjust the target dose homogeneity, the dose constraints and their strength on the critical organs that may have defined in the Clinical Protocol.

In order to perform crowd-knowledge-based planning benchmark of AutoMBM against other non-AP treatment planning systems (TPS), the same planning protocol, as defined in the original TROG planning competition, was adhered to by this study (Supplementary Table S1) except that all participants were demanded to achieve 20 Gy to cover 99% of every GTV as a hard constraint. Other plan quality metrics are also given in Table S1. All institutions applied their own AP templates to reflect their clinical practices. The general approach to achieve the optimal plan quality metrics among institutions that planned by AutoMBM v.1.5 was primarily by changing the AP template. Among institutions planning by AutoMBM v.2.0, the dose distribution could have been further optimized by adjusting the prescription isodose line

for individual targets, and the dose-volume constraints and their strength on other critical organs when necessary. Final doses were calculated pencil beam algorithm with adaptive dose calculation grid of resolution from 0.63 to 1.25  $\text{mm}^3$ , except for one institution that applied Monte Carlo dose-to-medium calculation at a resolution of  $1.9 \times 1.9 \times 2.0 \text{ mm}^3$  and statistical uncertainty of 2 %.

#### Extraction and analysis of plan quality metric

For all cases, the dose matrices were exported in DICOM format of uniform 1  $\text{mm}^3$  resolution. The objective scoring algorithm underlying the ProKnow platform was developed by Nelms et al. [21]. Each plan was scored based on 23 metrics including target coverage per lesion, Paddick conformity index (PCI) [22] and R50% [23] as a proxy to the steepness of dose gradient per total as well as individual lesion volume, etc (Supplementary Table S1).

Each metric has a maximum score and a baseline score for the ideal value and the minimum requirement, respectively. Zero score was given to metric that did not meet the minimum requirement. The sum of the scores of these 23 metrics was total to 150 points. All scoring metrics were obtained directly from the AutoMBM software and their values were populated onto an Excel sheet (Microsoft Excel version 2102, Microsoft Corporation, Redmond, USA) with formulae written specifically for this study to calculate their respective scores and the composite score according to the exact scoring functions devised in the TROG planning competition (Supplementary Table S1). It is important to note that the dosimetry scoring matrix was based on the Local HER 0 trial protocol [19] and did not necessarily reflect the clinical practice of individual participant.

For comparison, this study crowdsourced the population statistics of seven TPS (Table 1) that were employed in the TROG planning competition. One of the TPS, Pinnacle (Philips Radiation Oncology Systems, Fitchburg, WI, USA), was used in three plan submissions, its population statistics were crowdsourced but were not included in the statistical comparisons. Note that different techniques such as static-field/arc modulated radiotherapy (IMRT / VMAT), dynamic arc radiotherapy (DCA) delivered with circular collimator or multileaf collimator (MLC), and single or multiple isocenters may have been planned and were combined to produce the population statistics per TPS. The population statistics that are publicly accessible from the cloud-based ProKnow system are the medians, means, and one standard deviations (S.

**Table 1**  
Characteristics of the AutoMBM plan submissions in MC-PRIMA study.

	Number of institutions	% of total
<b>Nominal photon energy</b>		
6MV	15	62.5
6MV – flattening filter free	9	37.5
<b>MLC width (mm)</b>		
2.5 (inner 8 cm) and 5 (outer 14 cm); max field size $22 \times 30 \text{ cm}^2$	18	75.0
5.0 (central 20 cm) and 10 mm (outer 20 cm); max field size $40 \times 40 \text{ cm}^2$	2	16.7
5.0; max field size $40 \times 40 \text{ cm}^2$	4	8.3
<b>Number of couch positions</b>		
$\geq 3$	24	100.0
$>4$	22	91.7
$>5$	3	12.5
$>6$	1	4.2
$>7$	0	0.0
<b>Gantry arc length (°)</b>		
$\geq 100$	121	100.0
$\geq 120$	105	86.8
$\geq 140$	84	69.4
$\geq 160$	0	0.0

<sup>1</sup> <https://proknowsystems.com/>

D.) of the 23 plan quality metrics and their respective scores.

The PCI and dose gradient index (GI) [24] values per lesion that were calculated automatically by the AutoMBM plans were also recorded for quantitative analyses. Besides collection of the SRS plans, experiences of the planners were documented.

All centers calibrated the machine output so that one monitor unit gave one cGy at the depth of maximum dose for a reference  $10 \times 10 \text{ cm}^2$  field.

*Comparison of plan variability and quality between TPS*

To facilitate statistical comparisons of the variability and averaged performance of dose statistics between AutoMBM and each of the other TPS, this study assumed that the sample followed a normal distribution with sample number, mean and standard deviation known for each TPS. For comparison of the standard deviation between AutoMBM and each of the other TPS, the Levene’s tests were used. For comparison of mean, we used the two sample *t*-tests with either equal or unequal variances for PCI, R50%, GI, normal brain volume receiving 10 and 12 Gy ( $NB_{V10Gy}$  and  $NB_{V12Gy}$ ), maximum ( $D_{max}$ ) dose of the chiasm, eyes and lens, dose to  $0.3 \text{ cm}^3$  ( $D_{0.3cc}$ ) of the brainstem, volume receiving 8 Gy ( $V_{8Gy}$ ) in the optic nerves and mean dose ( $D_{mean}$ ) to the hippocampuses, the composite plan score and the monitor units (MU) as complexity metrics [25], according to the results of the preceding Levene’s tests. Normal distributions of different dose metrics were generated and statistical tests were performed using Matlab v. R2018a (Mathwork Inc. MA, USA).

*Dependence of AutoMBM plan quality on the planner’s experience and the nature of the treating center*

The potential of AutoMBM to lower or even eliminate the dependence of the plan quality in terms of composite plan score on the general and SRS planning experiences and the nature of SRS treating center, i.e., academic vs. non-academic, were evaluated by Pearson’s correlation and two sample *t*-test, respectively. Shapiro-Wilk tests for normality of quantities in all correlation analyses were performed.

**Results**

*Autoplanning with AutoMBM*

Table 1 summarizes the characteristics of the AutoMBM submission plans. The majority of institutions applied five table angles (19 of 24) and a gantry arc length of  $160^\circ$  in the AP. Fig. 1 summarizes the distributions of the treatment table angle and the gantry arc length defined by different planners in their AP templates, and the automatically optimized collimator angle.

All participants conformed to the requirement of 20 Gy covering 99% of five targets. The prescription isodose line normalized to the maximum dose ranges from 59 to 89% (mean  $\pm$  one 1 SD =  $80 \pm 6\%$ ). Fig. 2 plots the PCI (20 Gy) vs. GI for each GTV. Mean and SD of the PCI are  $0.58 \pm 0.05$ ,  $0.67 \pm 0.05$ ,  $0.55 \pm 0.09$ ,  $0.71 \pm 0.04$  and  $0.67 \pm 0.06$ , and for the GI  $4.41 \pm 0.61$ ,  $4.47 \pm 0.82$ ,  $6.55 \pm 1.23$ ,  $3.23 \pm 0.31$  and  $5.41 \pm 0.92$  for GTV 1 to 5, respectively. Fig. 3 shows the results of PCI, GI, R50% of all GTVs, and normal brain volume receiving 12 and 10 Gy ( $NB_{V10Gy}$  and  $NB_{V12Gy}$ ) of the AutoMBM plans from individual participants.

*Comparison of plan variability and quality between TPS*

Table 2 gives the dosimetric results of the AP solution using AutoMBM and non-AP solutions using other TPS that were crowd-sourced from the cloud-based ProKnow system. The statistical significance at  $p < 0.05$  in the comparison of the mean and the standard deviation is indicated by the bold value in Table 2. Supplementary Table S2 gives the corresponding objective score for each evaluated target and OAR, normalized to the respective maximum per TPS.

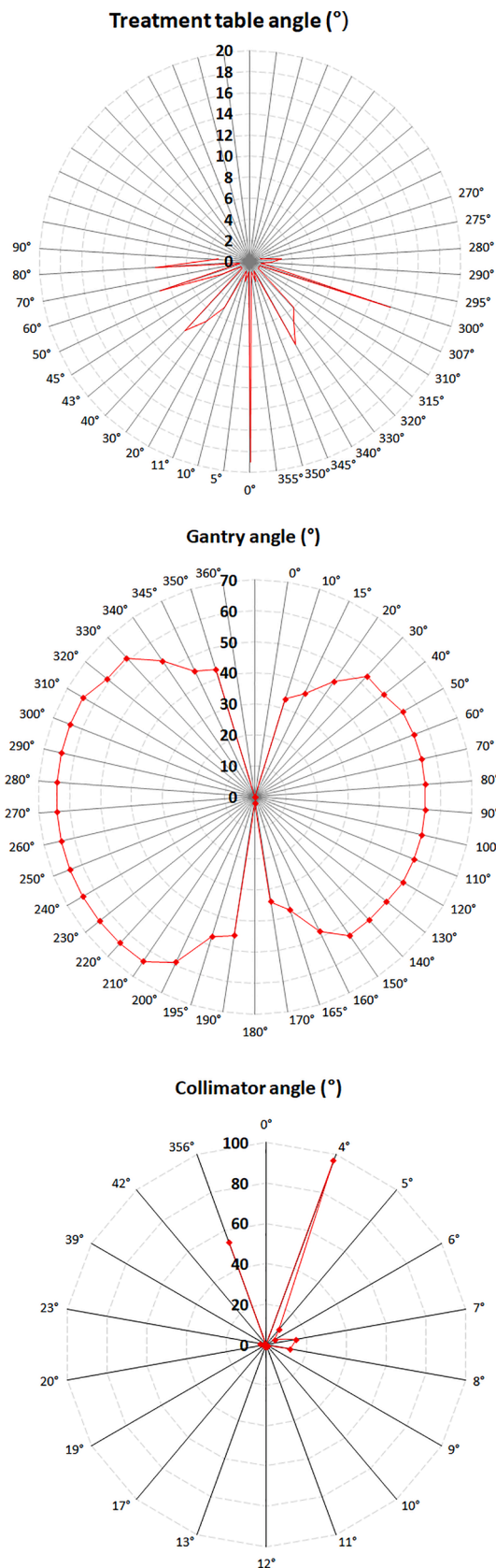


Fig. 1. Polar graphs showing the distributions of table, gantry angles covered the arcs and collimator angles for twenty-four AutoMBM plans. Each axis (solid grey) in these graphs represent the angle of the table, gantry and collimator and the radial grid lines (dotted light grey) indicate the number of these table, gantry and collimator angles, respectively.

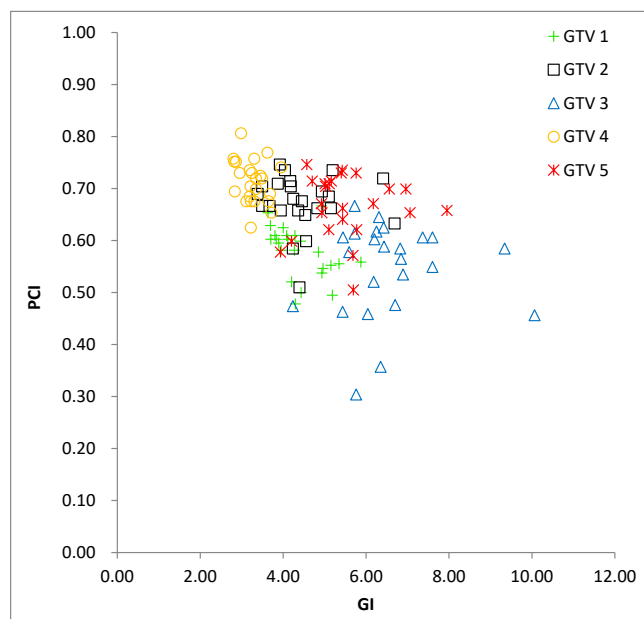


Fig. 2. Paddick Conformity index (PCI) vs. dose gradient index (GI) for individual gross tumor volumes (GTV) from twenty-four AutoMBM plans.

*Dependence of plan quality on the planner’s experience and the nature of the treating center*

The overall plan quality, evaluated by the composite score, shows no dependence of the participants’ experience in SRS and general planning (Fig. 4), with Pearson’s correlation coefficients  $r = -0.06$  ( $p = 0.767$ )

and  $-0.06$  ( $p = 0.764$ ), respectively. These statistical results are not affected after adjusting for the potentially influential parameter of the machine’s MLC width in the ANOVA tests. Furthermore, the difference in the means of the composite score between academic and non-academic centers is not significant (two sample  $t$ -test;  $p = 0.975$ ), and continues to be insignificant further adjusting for the technical factor of MLC width (ANOVA;  $p = 0.481$ ).

**Discussions**

The evaluation of automated stereotactic radiosurgery (SRS) planning against other manual solutions by either inverse or forward optimization is often subject to bias even in multi-center studies. This study overcame the classical limitation of planning benchmark by adapting a prior web-based plan challenge study from which dose metric statistics of 160 plans from a range of treatment planning systems (TPS) can be crowdsourced. This allowed us to bypass the need to involve many institutions to generate a large number of comparison plans yet enabling adequate statistics power in the critical appraisal of autoplanning by the Multiple Brain Mets SRS™ (AutoMBM) software. Another main advantage of crowdsourcing plan results from a heterogeneity of academic, non-academic and standalone institutions is that biases due to the variables of individual experiences in SRS and equipment specific characteristics (i.e., treatment machines) were effectively reduced.

*Characteristics of AutoMBM plans*

The AutoMBM incorporates the user-defined template in its AP solution. By such AP approach the planner reserves some degree of freedom to navigate the solution space for possible better plan dosimetry. Definition of these templates and hence the beam geometry is generally a non-trivial task, requiring certain level of expertise of the

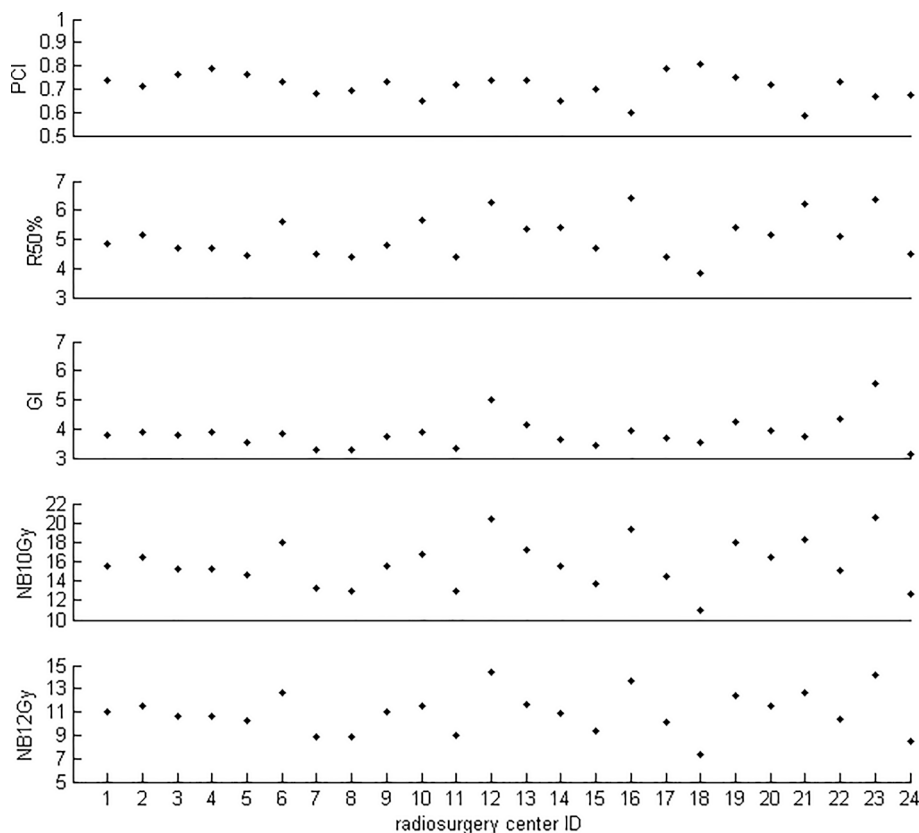


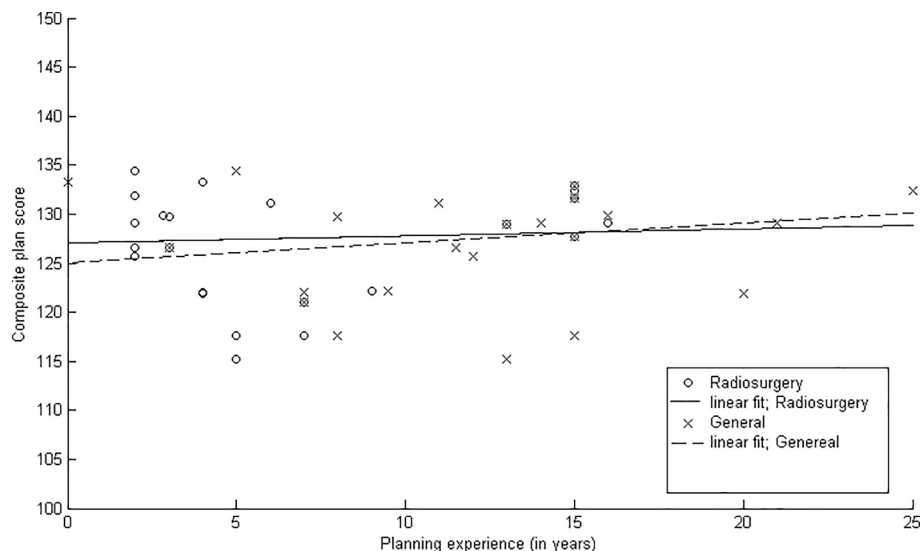
Fig. 3. Paddick conformity index (PCI), spread of half prescription isodose (R50%), dose gradient index (GI) of all gross tumor volumes, normal brain volume receiving 12 and 10 Gy ( $NB_{V12Gy}$  and  $NB_{V10Gy}$ , respectively) of the AutoMBM plans from individual participants.

**Table 2**

Means ± one standard deviations (SD) of various dosimetric parameters calculated for AutoMBM and crowdsourced for different planning solutions from the cloud-based ProKnow platform.

Dosimetrics	Treatment planning system							
	AutoMBM <sup>1</sup> (n = 24)	Eclipse <sup>2</sup> (n = 61)	Monaco <sup>3</sup> (n = 47)	RayStation <sup>4</sup> (n = 6)	iPlan <sup>4</sup> (n = 4)	MultiPlan <sup>6</sup> (n = 16)	GammaPlan <sup>7</sup> (n = 20)	Pinnacle <sup>8</sup> (n = 3)
<b>Gross tumor volume (GTV)</b>								
GTV1 V20Gy (%)	99.0 ± 0	98.8 ± 1.1	<b>99.4 ± 0.5</b>	99.4 ± 0.5	<b>99.7 ± 0.3</b>	99.2 ± 0.5	<b>98.1 ± 1.5</b>	99.7 ± 0.2
GTV2 V20Gy (%)	99.0 ± 0	98.9 ± 1.0	<b>99.4 ± 0.6</b>	99.7 ± 0.3	99.7 ± 0.4	99.2 ± 0.7	<b>98.1 ± 1.6</b>	99.2 ± 0.3
GTV3 V20Gy (%)	99.0 ± 0	98.9 ± 1.4	<b>99.5 ± 0.6</b>	<b>99.6 ± 0.2</b>	<b>99.6 ± 0.3</b>	<b>98.2 ± 2.3</b>	98.3 ± 1.5	99.3 ± 0.3
GTV4 V20Gy (%)	99.0 ± 0	98.9 ± 1.4	99.1 ± 0.7	99.4 ± 0.7	81.1 ± <b>32.1</b>	<b>99.5 ± 0.4</b>	98.7 ± 0.8	99.3 ± 0.2
GTV5 V20Gy (%)	99.0 ± 0	98.8 ± 1.4	96.8 ± 0.8	99.0 ± 0.9	99.4 ± 0.4	98.5 ± 1.3	<b>97.5 ± 2.9</b>	99.4 ± 0.4
PCI of all GTVs(20 Gy)	0.71 ± 0.06	0.73 ± 0.09	<b>0.64 ± 0.12</b>	<b>0.64 ± 0.02</b>	<b>0.54 ± 0.20</b>	0.71 ± 0.06	0.69 ± 0.06	0.51 ± 0.10
R50% of all GTVs(10 Gy)	5.01 ± 0.73	5.48 ± 1.68	<b>8.49 ± 6.60</b>	<b>9.70 ± 3.46</b>	4.76 ± 0.90	4.65 ± 0.82	<b>4.06 ± 0.44</b>	11.85 ± 2.14
<b>Normal organs</b>								
Normal brainV12Gy (cc)	10.7 ± 1.9	10.9 ± 4.0	<b>18.3 ± 15.5</b>	<b>19.0 ± 31.8</b>	10.2 ± 2.1	9.4 ± 1.8	8.1 ± 1.2	10.2 ± 2.1
Normal brainV10Gy (cc)	15.8 ± 2.5	16.6 ± 6.0	<b>27.3 ± 23.0</b>	<b>31.8 ± 12.7</b>	14.4 ± 2.8	13.5 ± 2.7	11.7 ± 1.7	37.8 ± 5.8
Right hippocampus Dmean (Gy)	2.4 ± 0.6	<b>1.8 ± 0.9</b>	<b>1.9 ± 1.2</b>	2.3 ± 1.1	2.2 ± 1.1	1.8 ± 0.8	<b>1.9 ± 0.2</b>	2.0 ± 0.33
Left hippocampus Dmean (Gy)	1.8 ± 0.3	<b>1.3 ± 0.6</b>	1.5 ± 1.7	0.6 ± 1.3	<b>0.5 ± 0.5</b>	1.3 ± 0.5	<b>1.0 ± 0.1</b>	1.4 ± 0.2
Right eye Dmax (Gy)	3.8 ± 1.2	4.0 ± 1.4	4.2 ± 1.7	5.3 ± 2.3	2.8 ± 1.0	3.1 ± 2.2	3.2 ± 0.4	4.5 ± 0.7
Left eye Dmax (Gy)	4.0 ± 1.2	4.0 ± 1.2	4.2 ± 1.7	5.2 ± 1.6	3.0 ± 0.8	3.2 ± 2.7	3.6 ± 0.6	3.9 ± 0.4
Right lensDmax (Gy)	1.8 ± 0.5	1.5 ± 0.6	1.6 ± 0.6	1.7 ± 0.2	<b>0.9 ± 0.2</b>	1.0 ± 0.8	1.6 ± 0.2	1.4 ± 0.2
Left lensDmax (Gy)	1.6 ± 0.4	<b>1.3 ± 0.5</b>	1.4 ± 0.5	1.6 ± 0.3	1.1 ± 0.3	1.1 ± 1.0	<b>1.4 ± 0.3</b>	1.4 ± 0.4
Right optic nerveV8Gy (cc)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Leftoptic nerveV8Gy (cc)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Chiasm Dmax(Gy)	4.5 ± 1.1	<b>3.7 ± 1.2</b>	<b>3.2 ± 1.2</b>	<b>0.9 ± 0.3</b>	3.2 ± 0.7	<b>2.9 ± 0.9</b>	<b>3.6 ± 0.8</b>	3.5 ± 1.2
Brainstem D0.3 cm <sup>3</sup> (Gy)	5.8 ± 0.8	6.4 ± 1.7	<b>7.3 ± 2.6</b>	<b>9.1 ± 1.8</b>	5.3 ± 1.1	5.1 ± 0.9	<b>4.9 ± 0.8</b>	10.1 ± 0.9
Composite plan score	127.6 ± 6.2	127.3 ± 14.7	<b>118.3 ± 15.0</b>	<b>108.6 ± 10.3</b>	126.7 ± 7.3	131.7 ± 12.6	127.3 ± 11.4	101.7 ± 4.1
Monitor units	8823.6 ± 4154.3	<b>17,703 ± 9999.6</b>	<b>17679.5 ± 7895.55</b>	<b>14434.1 ± 13162.5</b>	8071.0 ± 7744.1			14,261 ± 4415.7

<sup>1</sup> AutoMBM (Brainlab AG, Feldkirchen, Germany); <sup>2</sup>Eclipse (Varian Inc. Palo Alto, CA, USA); <sup>3</sup>Monaco (Elekta Oncology System, Crawley, UK); <sup>4</sup>RayStation (Raysearch Laboratories, Stockholm, Sweden); <sup>5</sup>iPlan (Brainlab AG, Feldkirchen, Germany); <sup>6</sup>MultiPlan (Accuray Inc., Sunnyvale, CA, USA); <sup>7</sup>GammaPlan (Elekta AB, Stockholm, Sweden); <sup>8</sup>Pinnacle (Philips Radiation Oncology Systems, Fitchburg, WI, USA). Note: Bold values indicate statistical significance ( $p < 0.05$ ) between MBM<sup>1</sup> and each of the other TPS<sup>2-8</sup>.



**Fig. 4.** Plot of radiosurgery and general planning experiences versus composite plan score of AutoMBM.

individual planner as in conventional planning. Fig. 1 depicts distinctive preference among individual planners to use couch angle at 0° and 330° and to limit the couch angle to within 40° from 0°. The vast majority of planners appeared to practice the similar planning philosophy that avoided shoot-through the longitudinal body axis at either 90° or 270° couch angles. Fig. 1b also depicts the preference of most planners to use

half gantry arcs and to omit the first 10°–20° from either 0° or 180° respectively to avoid direct opposing beams. On the other hand, the algorithm of AutoMBM overwhelmingly favored small collimator rotations, viz, 4°, although larger rotations over 30° were also observed. It is worth noting that this study did not collect individual planner’s AP templates and therefore was unable to analyze the minimum collimator



angle defined therein. Nevertheless, most of the planners was believed to have set the minimum collimator as  $4^\circ$  that was favourably chosen by the AutoMBM. The auto-optimized collimator angle may depend on a number of factors such as the couch angle, gantry arc span as well as the machine characteristics (e.g., maximum field size and width of multi-leaf collimator (MLC), etc). For all plans that resulted in  $> 10^\circ$  collimator rotation, the collimator aperture had a limited field size in one dimension of 22 cm.

#### *Inter-planner variation in plan quality using AutoMBM and other non-Auto TPS*

The heterogeneity of the template definition and machine characteristics did not appear to contribute to excessive variability in the plan quality from AutoMBM. Considering two of the most concerning metrics in SRS, the values of PCI and GI showed relatively small dispersion among the institutions in each lesion (Fig. 2). Large GI (10.06 and 9.34) and low PCI (0.30 and 0.36) all corresponded to the smallest lesion of  $0.07 \text{ cm}^3$ . Interestingly, the machine characteristics, and more specifically concerning the width of the MLC, might not be the absolute factor attributing to these outlying values as it seemed as there were other AutoMBM plans created for the same type of machine and 5 mm MLC on Elekta Agility (Elekta, Crawley, UK) and Varian Millennium 120 (Varian, CA, USA) achieving GI of 6.89 and 6.84, and PCI of 0.53 and 0.56 for this smallest lesion, respectively. Furthermore, the beam arc geometry of these outlying plans did not differ substantially from other AutoMBM plans either, regardless of the machine / MLC model. What may influence the plan quality remains in the different settings of the Clinical Protocol template and the level of planner-enabled interactive smart tuning for controlling the degree of normal tissue sparing and MU spread during the optimization.

An important aspect that influenced the variability of plan quality among planners from different institutions is the treatment planning system (TPS) in use. This study showed that AutoMBM did not differ from most other TPS without using SRS-dedicated AP concerning the variability of PCI, except for Monaco (Elekta, MO, USA) and iPlan (BrainLab, Munich, Germany). There were, however, marked differences in the inter-planner variability of R50% among TPS. AutoMBM reduced this variability compared to some TPS, especially those that are not dedicated to SRS such as Eclipse (Varian, CA, USA), Monaco and RayStation (Raysearch, Stockholm, Sweden). The fact that the GammaPlan TPS dedicated to GammaKnife SRS (Elekta, Stockholm, Sweden) achieved significantly smaller inter-planner variability in R50% compared to AutoMBM was likely associated with the historical practice of dose prescription at about the 50% isodose [26]. We also observed such practice in most participants using GammaPlan whereas the distribution of prescription isodose level was much wider for planners using Eclipse, Monaco, and RayStation in the TROG planning competition. The difficulty to achieve uniform PCI and GI using Monaco was found even within the same institution. Hofmaier et al. obtained a wide range of CI and GI values from 0.38–0.88 and 3.35–33.0 for the Monaco plans, which were reduced to 0.58–0.89 and 3.50–15.73 by AutoMBM [11]. Similar difficulty to obtain uniform GI within the same institution was reported for Eclipse [11]. Gevaert et al. achieved GI at one standard deviation (1 S.D.) of 3.1 vs. 1.6 planning VMAT on Eclipse without AP and dynamic conformal arc (DCA) by AutoMBM, respectively [12]. On the other hand, the majority of planners using iPlan were believed to have performed forward dose optimization for DAC delivery using circular collimator or MLC. Such approach is effective at producing good target dose conformity when the target shape is fairly regular [27]. Given the poor sphericity of the targets in this TROG benchmark case, the possibility of creating a complex dose distribution confirming to the irregular targets' surfaces became critically planner-dependent and hence significant variability of the PCI.

One may anticipate lower variability of plan quality when plans were generated for treatment platforms from the same vendor. This was

generally true for the dedicated SRS delivery platforms but not for the linac-based platforms. The one S.D. of composite plan score is comparable between MultiPlan for CyberKnife (Accuray, CA, USA), GammaPlan for GammaKnife and AutoMBM. For CyberKnife planning on other vendor-independent TPS which has recently become possible on RayStation, the proposed crowdsourcing approach is deemed to be useful to efficiently evaluate the inter-institution / planner variability compared with the vendor-dependent MultiPlan. For plans that were generated on Monaco and Eclipse for linacs presumably belonging to the same vendors (Elekta and Varian, respectively), the excessive variability of composite plan scores ( $p < 0.05$ ) compared to AutoMBM could be partly attributed to the absence of dedicated AP algorithm to manage the overlapping of MLC apertures between targets and the sharing of same pairs of MLC by two or more targets that increased the dose bridging between targets in the normal brain. This problem also applied to the other TPS independent of the linac vendor such as RayStation. As demonstrated in previous studies [18,28,29], the dose bridging problem cannot be easily resolved even with collimator angle optimization and finer MLC width owing to the intrinsic limitation of the optimizer in couple with MLC sequencer. In contrast, AutoMBM automated the allocation of targets among arcs to treat as many targets by as many arcs as possible by optimizing the collimator angle while avoiding two or more metastases sharing the same pair of MLC. If two targets shared a leaf pair, the targets were assigned to different arcs at the same couch position. Else, both targets were assigned to the same arc. This fully automated planning process was responsible for reducing the planner's intervention in the control of the intermediate-to-low dose spill and consequently reduced the variability of R50%, and the normal brain receiving 10 and 12 Gy (NB<sub>V10Gy</sub> and NB<sub>V12Gy</sub>, respectively).

The uniqueness of AutoMBM to achieve uniform target dose coverage was clearly demonstrated in Table 2. For one reason, this study demanded for every lesion that the planner must prescribe 20 Gy to cover 99% of the GTV (GTV  $V_{20\text{Gy}}$ ). The algorithm of AutoMBM guaranteed the precise prescription of GTV  $V_{20\text{Gy}} = 99\%$  by a stochastic optimization of arc weights that followed the final dose distribution resulting from the optimized arc configuration and dynamic arc MLC sequencing. With other TPS, renormalization was inevitably needed given deviations of the desired target coverage at the end of the final dose calculation. This process involved non-trivial trade-off between target coverage and PCI that varied according to the clinical preference of individual planners and radiation oncologist. When separate plans were created for different lesions, the optimal target coverage vs. PCI that had been achieved by renormalization for one plan was likely to change after plans for individual lesions were summed. This situation necessitated repeated renormalization for every lesion by trial and error until the planner achieved the global but also compromised optimality of target coverage and PCI for all lesions. In case where all lesions were co-optimized in one single plan, the renormalization after the final dose calculation would affect the target coverage and PCI of individual lesions all in once, making it nearly impossible to achieve uniform target coverage.

As the overall effect, AutoMBM achieved significantly more uniform composite plan score compared to other non-AP TPS for linac-based SRS, as evidenced in Table 2. Moreover, the larger variability of plan quality also suggested greater spread of plan complexity which was inferred by the significantly larger number of monitor units [30].

#### *Overall plan quality between AutoMBM and other non-autoplan TPS*

The ability of harmonizing the SRS plan quality is by far not enough to consider AutoMBM as viable solution to SRS for multiple brain metastases. Equally important is that AutoMBM achieves plan quality standard that is comparable to other non-AP TPS. Despite the statistical significance,  $V_{20\text{Gy}}$  of GTV1-5 was grossly comparable across different TPS regardless of AP. One exception was observed with iPlan where the mean  $V_{20\text{Gy}}$  for GTV4 was merely 81% and was likely caused by an

extreme outlier as indicated by the one S.D. of 31%.

The fact that MBM could not achieve comparable dose fall R50%, NBV<sub>12Gy</sub> and NBV<sub>10Gy</sub> to dedicated radiosurgery TPS GammaPlan and MultiPlan despite AP augmentation was not entirely surprising. It was largely because AutoMBM was a AP solution to linac-based radiosurgery using MLC that shows broader dosimetric penumbra compared with conical collimators used in GammaKnife and CyberKnife. GammaPlan and MultiPlan also used a large number of isocenters and non-isocenter confocal beams which could significantly protect the normal brain and other OARs than other TPS for non-coplanar radiosurgery on linacs [31–34]. Nonetheless, the composite plan scores of GammaPlan, MultiPlan and AutoMBM were statistically equal although AutoMBM showed slightly worse dose statistics in other OARs in general. Because the scoring functions (Supplementary Table S1) that were designed per the Local HER-0 trial had higher weights on the target coverage, AutoMBM scored more points from the ideal target coverage for all lesions than GammaPlan and eventually equivalent composite plan score. On the other hand, MultiPlan achieved almost full scores from target coverage, and higher scores from better CI, NBV<sub>12Gy</sub> and other OARs and eventually higher composite plan score despite statistical insignificance. Rossi et al. have recently shown that the boundary of the CyberKnife plan quality could be pushed even further by AP using the vendor-independent iCycle software in prostate radiotherapy [35]. As another vendor-independent TPS has also become available for CyberKnife planning, further crowdsourcing plan benchmark will provide more insights into the role of TPS and the incorporation of AP in the overall plan quality.

Results for linac-based SRS were different between TPS, with Eclipse and iPlan showing very similar R50%, normal brain doses and doses in other OARs to AutoMBM, while Monaco and RayStation almost reversed the results in comparison to AutoMBM. Although Monaco and RayStation scored almost the maximum possible points of 45.9 and 49.9 out of 50 for GTV V<sub>20Gy</sub>, the worse performance in PCI, R50%, NBV<sub>10Gy</sub>, NBV<sub>12Gy</sub> and dose to 0.3 cc of brainstem (D<sub>0.3cc</sub>) rendered them to lose a great deal of points to these high-weighted dosimetric metrics and ultimately significantly lower composite plan scores. A possible reason for these results could be that a large number of plans on Eclipse and iPlan were created for Varian linacs assuming the finest MLC width of 2.5 mm while Monaco for Elekta linacs assuming the finest MLC width of 4–5 mm [28,36,37]. Regardless of this hypothesis, it is still clear that planners using AP-powered were able to overcome the influence of machine / MLC configuration and achieve comparable and even superior dose statistics to other existing TPS. The ability of AP to reduce the plan quality variability among treatment machines is not unique to AutoMBM but also reported for other TPS in other treatment sites of head and neck, pancreas and rectal cancers [38].

### Limitations

The dosimetric scoring matrix and the eventual composite plan score were uniquely devised based on Local HER-0 trial protocol. There was other cloud-based international plan challenge study that used different scoring matrix [18]. Although the interpretation of overall plan quality may change with the scoring algorithm and the definition of the plan quality [39], the results that AutoMBM could reduce the inter-planner variability and improve the averaged performance of a range of dose metrics remain valid. The retrospective nature of this study precluded the case selection bias in favor of AutoMBM. Nevertheless, there was a rare possibility of bias stemming from the prior knowledge of scores by the other TPS which might steer the participants to work strategically on certain scoring metrics to higher scores. Yet, this study achieved RAT-ING scores of 94% and 92% by two authors (XX and XX) [40].

It was acknowledged that the statistical comparisons may be subject to debate because of the assumption of normality in the scoring metrics. The ProKnow system provided only the population statistics of dosimetric results per either TPS or delivery technique. Statistical

comparisons using the Levene's tests for variance and the two sample *t*-tests for mean represented the best effort and were deemed reasonable for quantifying the relative performance of TPS. As this study focused on TPS as the main factor that influenced the plan quality and variability, the impact of delivery technique was ignored, which was partly due to lack of the information about the delivery techniques that were planned per TPS from the ProKnow system. Delivery technique is known to influence the plan quality. For example, dosimetric comparisons showed PCI and GI on the same TPS with VMAT better than IMRT [41] and comparable between mono-and multi-isocentric DCA [12]. Contradictory results also existed comparing for the same delivery technique on the same TPS, for example, cone-based vs. MLC-based DCA on linac [8,42,43] and on CyberKnife [44,45]. Although institutions shall not be limited with the choice of delivery technique, the above studies together with the present MC-PRIMA suggest especially for linac-based SRS to multiple brain metastases that the delivery technique(s) shall be carefully chosen per TPS. We strongly recommend that organizers of international plan competitions share detailed information of individual plan submission on the web-platforms for their upcoming as well as past studies such as the one for the TROG Local HER 0 trial study case. The availability of this information is important to fully unlock the potential of plan crowdsourcing when potentially interacting factors such as TPS, machine characteristics and delivery technique could be taken into account in the statistical analyses [46]. Upon the ultimate goal of knowledge sharing via cloud based plan crowdsourcing, SRS institutions can be more informed of which TPS combined with what delivery technique to avoid suboptimal and large variability of plan quality.

The plan complexity has an implication on the plan deliverability [25]. In the original TROG plan competition and this study involving a spectrum of delivery techniques planned by different TPS, the number of MU per Gy may be the most simplest and applicable complexity metrics for evaluation and comparison of the relative plan deliverability between TPS [47]. Other complexity metrics, such as modulation complexity score [48] and variations of the nominal dose rate or gantry speed [49], etc, suffer from limited application to certain delivery technique that could be planned by certain TPS and therefore unsuitable for this study. In Table 2, the MU resulting from AutoMBM was found to be significantly lower than from other TPS, which may result in superior deliverability [30]. Although highly recommended [40], a dry-run test for the deliverability and the dosimetric accuracy of the AutoMBM submission plan was not requested in MC-PRIMA. It is common that the post-planning dosimetric validation was left out in non-sponsored multi-center planning studies [10,13,46,50], like MC-PRIMA, predominately because financial resource was generally beyond reach to arrange rigorous dosimetry audit like in clinical trials. Nonetheless, all institutions that participated in MC-PRIMA had rigorous commissioning program and routine plan quality assurance in place to demonstrate acceptable deliverability and accuracy of their clinical AutoMBM plans.

### Dependence of plan quality on the planner's experience and the nature of treating center

This study collected information about the planning experience of each individual planner. When the composite plan score was plotted versus the general planning and radiosurgery planning experience of the planner, no significant correlations were found. Plan scores resulting from those planners from academic were also compared with the others from non-academic centers, again without observed significant difference. Other studies [51,52] have showed that AutoMBM could outperform human planners but this study, to our best knowledge, is the first to demonstrate that AutoMBM could also eliminate the dependence of plan quality on the planning experiences and the nature of the treating center. When radiosurgery is increasingly applied to treat multiple brain metastases, AutoMBM proves to offer a viable option to alleviate the problem of global shortage of planners / dosimetrists that typically take years to develop their expert skills. Further crowdsourcing plan

benchmark from other cloud-based international plan challenge studies is warranted to validate the results of this MC-PRIMA study and to investigate that other AP in general are able to improve plan quality and variability in SRS planning for multiple brain metastases.

## Conclusions

Plan crowdsourcing offers an efficient means to benchmark new TPS or delivery technique. This plan crowdsourcing study shows promises of AutoMBM to achieve clinically acceptable plans with minimal inter-planner variability for SRS to multiple intracranial metastases independent of the planning experience and the institution.

## Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: M.K.H. Chan was an employee at Imperial College NHS Healthcare Trust, UK and University Hospital Essen, Germany at the time of preparing of this study. Current study was conducted without involvement of these employments. Matthew Wong was an employee at Tuen Mun Hospital, China at the time of participation in this study. The authors received technical support in the collection of plan submissions and technical information of the Multiple Brain Metastases SRSTM software from Brainlab (Brainlab AG, Munich, Germany). No authors received financial support involved in the study design or materials, analysis and interpretation of data nor in the writing of the publication. Part of this work was presented as e-Poster in ASTRO annual meeting 2019.

## Acknowledgements

The authors would like to thank all participants who contributed to the plan submissions, including Philippa Sturt at Royal Marsden Hospital NHS Foundation Trust, UK, Sylvia Garry at University Hospital, Ludwig-Maximilians University of Munich, Germany, Manfred Wienzier at University Hospital Essen, Germany, Daniel Venencia at Instituto Zunino-Fundacion Marie Curie, Argentina, Alexander Malesevich at Brainlab, Germany, Jérôme Krayenbuehl at University Hospital Zürich, Switzerland.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejmp.2022.01.011>.

## References

- Churilla TM, Ballman KV, Brown PD, Twohy EL, Jaeckle K, Farace E, et al. Stereotactic radiosurgery with or without whole-brain radiation therapy for limited brain metastases: a secondary analysis of the north central cancer treatment group N0574 (Alliance) randomized controlled trial. *Int J Radiat Oncol Biol Phys* 2017;99(5):1173–8.
- Chang WS, Kim HY, Chang JW, Park YG, Chang JH. Analysis of radiosurgical results in patients with brain metastases according to the number of brain lesions: is stereotactic radiosurgery effective for multiple brain metastases? *Clinical article J Neurosurg* 2010;113(Special Supplement):73–8.
- Palmer JD, Sebastian NT, Chu J, DiCostanzo D, Bell EH, Grecula J, et al. Single-isocenter multitarget stereotactic radiosurgery is safe and effective in the treatment of multiple brain metastases. *Pract Radiat Oncol* 2020;5(1):70–6.
- Raldow AC, Chiang VL, Knisely JP, Yu JB. Survival and intracranial control of patients with 5 or more brain metastases treated with gamma knife stereotactic radiosurgery. *Am J Clin Oncol* 2013;36(5):486–90.
- Yamamoto M, Serizawa T, Shuto T, Akabane A, Higuchi Y, Kawagishi J, et al. Stereotactic radiosurgery for patients with multiple brain metastases (JLKG0901): a multi-institutional prospective observational study. *Lancet Oncol* 2014;15(4):387–95.
- Eaton DJ, Lee J, Paddick I. Stereotactic radiosurgery for multiple brain metastases: results of multicenter benchmark planning studies. *Pract Radiat Oncol* 2018;8(4):e212–20.
- Foy JJ, Marsh R, Ten Haken RK, Younge KC, Schipper M, Sun Y, et al. An analysis of knowledge-based planning for stereotactic body radiation therapy of the spine. *Pract Radiat Oncol* 2017;7(5):e355–60.
- Krayenbuehl J, Zamburlini M, Ghandour S, Pachoud M, Tanadini-Lang S, Tol J, et al. Planning comparison of five automated treatment planning solutions for locally advanced head and neck cancer. *Radiat Oncol* 2018;13(1). <https://doi.org/10.1186/s13014-018-1113-z>.
- Ueda Y, Fukunaga J-I, Kamima T, Adachi Y, Nakamatsu K, Monzen H. Evaluation of multiple institutions' models for knowledge-based planning of volumetric modulated arc therapy (VMAT) for prostate cancer. *Radiat Oncol* 2018;13(1). <https://doi.org/10.1186/s13014-018-0994-1>.
- Liu H, Thomas EM, Li J, Yu Y, Andrews D, Markert JM, et al. Interinstitutional plan quality assessment of 2 linac-based, single-isocenter, multiple metastasis radiosurgery techniques. *Adv Radiat Oncol* 2020;5(5):1051–60.
- Hofmaier J, Bodensohn R, Garry S, Hadi I, Fleischmann DF, Eder M, et al. Single isocenter stereotactic radiosurgery for patients with multiple brain metastases: dosimetric comparison of VMAT and a dedicated DCAT planning tool. *Radiat Oncol* 2019;14(1). <https://doi.org/10.1186/s13014-019-1315-z>.
- Evaert T, Steenbeke F, Pellegrini L, Engels B, Christian N, Hoornaert M-T, et al. Evaluation of a dedicated brain metastases treatment planning optimization for radiosurgery: a new treatment paradigm? *Radiat Oncol* 2016;11(1). <https://doi.org/10.1186/s13014-016-0593-y>.
- Vergalasova I, Liu H, Alonso-Basanta M, Dong L, Li J, Nie KE, et al. Multi-institutional dosimetric evaluation of modern day stereotactic radiosurgery (SRS) treatment options for multiple brain metastases. *Front Oncol* 2019;9. <https://doi.org/10.3389/fonc.2019.00483>.
- Kadoya N, Abe Y, Kajikawa T, Ito K, Yamamoto T, Umezawa R, et al. Automated noncoplanar treatment planning strategy in stereotactic radiosurgery of multiple cranial metastases: HyperArc and CyberKnife dose distributions. *Med Dosim* 2019;44(4):394–400.
- Moghanaki D, Slotman B, Swaminath A, Nelms B, Wang B. Assessing the variability and quality of lung stereotactic radiation therapy treatment plans using a web-based crowdsourcing platform. *Pract Radiat Oncol* 2020;10(3):e118–27.
- Hardcastle N, Bignell F, Nelms B, Siva S, Kneebone A, Lao L, et al. The challenge of planning vertebral body SBRT: Optimizing target volume coverage. *Med Dosim* 2020;45(3):302–7.
- Hardcastle N, Nelms B, O'Connor L, Shakeshaft J, Haworth A, Cook O, et al. OC-0521 SRS plan quality with variation in modality: Results of an international planning competition. *Radiother Oncol* 2019;133:S272–3.
- Giles WM, Cullom ET, Laryea O-O, Nobah A, Alves VGL, Yin F-F, et al. Radiosurgery treatment planning using conformal arc informed volumetric modulated arc therapy. *Med Dosim* 2021;46(1):3–12. <https://clinicaltrials.gov/ct2/show/NCT02898727> [Accessed date: 21-06-2021].
- Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* 2012;30(9):1323–41.
- Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. *Pract Radiat Oncol* 2012;2(4):296–305.
- Paddick I. A simple scoring ratio to index the conformity of radiosurgical treatment plans. *J Neurosurg (Suppl 3)* 2000;93:219–22.
- Timmerman R, Paulus R, Galvin J, Michalski J, Straube W, Bradley J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA* 2010;303:1070–6.
- Paddick I, Lippitz B. A simple dose gradient measurement tool to complement the conformity index. *J Neurosurg* 2006;105(Supplement):194–201.
- McGarry CK, Agnew CE, Hussein M, Tsang Y, McWilliam A, Hounsell AR, et al. The role of complexity metrics in a multi-institutional dosimetry audit of VMAT. *Br J Radiol* 2016;89(1057):20150445. <https://doi.org/10.1259/bjr.20150445>.
- Dimitriadis A, Kirkby KJ, Nisbet A, Clark CH. Current status of cranial stereotactic radiosurgery in the UK. *Br J Radiol* 2016;89(1058):20150452. <https://doi.org/10.1259/bjr.20150452>.
- Lee YC, Kim Y. Circular collimator arc versus dynamic conformal arc treatment planning for linac-based stereotactic radiosurgery of an intracranial small single lesion: a perspective of lesion asymmetry. *Radiat Oncol* 2019;14:91.
- Ruschin M, Sahgal A, Iradji S, Soliman H, Leavens C, Lee Y. Investigation of two linear accelerator head designs for treating brain metastases with hypofractionated volumetric-modulated arc radiotherapy. *Br J Radiol* 2016;89(1063):20160093. <https://doi.org/10.1259/bjr.20160093>.
- Ohira S, Sagawa T, Ueda Y, Inui S, Masaoka A, Akino Y, et al. Effect of collimator angle on HyperArc stereotactic radiosurgery planning for single and multiple brain metastases. *Med Dosim* 2020;45(1):85–91.
- Chiavassa S, Bessieres I, Edouard M, Mathot M, Moignier A. Complexity metrics for IMRT and VMAT plans: a review of current literature and applications. *Br J Radiol* 2019;92(1102):20190270. <https://doi.org/10.1259/bjr.20190270>.
- Chea M, Fezzani K, Jacob J, Cuttat M, Croisé M, Simon J-M, 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. *Radiat Oncol* 2021;16(1). <https://doi.org/10.1186/s13014-021-01766-w>.
- Dong P, Pérez-Andújar A, Pinnaduwaage D, Braunstein S, Theodosopoulos P, McDermott M, et al. Dosimetric characterization of hypofractionated Gamma Knife radiosurgery of large or complex brain tumors versus linear accelerator-based treatments. *J Neurosurg* 2016;125(Supplement 1):97–103.
- Han EY, Wang H, Luo D, Li J, Wang X. Dosimetric comparison of fractionated radiosurgery plans using frameless Gamma Knife ICON and CyberKnife systems with linear accelerator-based radiosurgery plans for multiple large brain metastases. *J Neurosurg* 2020;132:1473–9.

- [34] Zhang S, Yang R, Wang X. Dosimetric quality and delivery efficiency of robotic radiosurgery for brain metastases: Comparison with C-arm linear accelerator based plans. *J Appl Clin Med Phys* 2019;20(11):104–10.
- [35] Rossi L, Sharfo AW, Aluwini S, Dirx M, Breedveld S, Heijmen B. First fully automated planning solution for robotic radiosurgery – comparison with automatically planned volumetric arc therapy for prostate cancer. *Acta Oncol* 2018;57(11):1490–8.
- [36] Chang J, Yenice KM, Jiang K, Hunt M, Narayana A. Effect of MLC leaf width and PTV margin on the treatment planning of intensity-modulated stereotactic radiosurgery (IMSRS) or radiotherapy (IMSRT). *Med Dosim* 2009;34(2):110–6.
- [37] Serna A, Puchades V, Mata F, Ramos D, Alcaraz M. Influence of multi-leaf collimator leaf width in radiosurgery via volumetric modulated arc therapy and 3D dynamic conformal arc therapy. *Phys Med* 2015;31(3):293–6.
- [38] Hirashima H, Nakamura M, Mukumoto N, Ashida R, Fujii K, Nakamura K, et al. Reducing variability among treatment machines using knowledge-based planning for head and neck, pancreatic, and rectal cancer. *J Appl Clin Med Phys* 2021;22(7):245–54.
- [39] Hernandez V, Hansen CR, Widesott L, Bäck A, Canters R, Fusella M, 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.
- [40] Hansen CR, Crijns W, Hussein M, Rossi L, Gallego P, Verbakel W, et al. Radiotherapy treatment planning study guidelines (RATING): a framework for setting up and reporting on scientific treatment planning studies. *Radiother Oncol* 2020;153:67–78.
- [41] Zhang S, Yang R, Shi C, Li J, Zhuang H, Tian S, et al. Noncoplanar VMAT for Brain Metastases: A Plan Quality and Delivery Efficiency Comparison With Coplanar VMAT, IMRT, and CyberKnife. *Technol Cancer Res Treat*. 2019;18:1533033819871621.
- [42] Shiu AS, Kooy HM, Ewton JR, Tung SS, Wong J, Antes K, et al. Comparison of miniature multileaf collimation (MMLC) with circular collimation for stereotactic treatment. *Int J Radiat Oncol Biol Phys* 1997;37(3):679–88.
- [43] Hsu S-M, Lai Y-C, Jeng C-C, Tseng C-Y. Dosimetric comparison of different treatment modalities for stereotactic radiotherapy. *Radiat Oncol* 2017;12:155.
- [44] Jang SY, Lalonde R, Ozhasoglu C, Burton S, Heron D, Huq MS. Dosimetric comparison between cone/Iris-based and InCise MLC-based CyberKnife plans for single and multiple brain metastases. *J Appl Clin Med Phys* 2016;17(5):184–99.
- [45] Murai T, Hattori Y, Sugie C, Iwata H, Iwabuchi M, Shibamoto Y. Comparison of multileaf collimator and conventional circular collimator systems in Cyberknife stereotactic radiotherapy. *J Radiat Res* 2017;58(5):693–700.
- [46] Esposito M, Masi L, Zani M, Doro R, Fedele D, Garibaldi C, et al. SBRT planning for spinal metastasis: indications from a large multicentric study SBRT-Planung bei Wirbelsäulenmetastasen: Indikationen aus einer großen Multizenterstudie. *Strahlenther Onkol* 2019;195(3):226–35.
- [47] Crowe SB, Kairn T, Middlebrook N, Sutherland B, Hill B, Kenny J, et al. Examination of the properties of IMRT and VMAT beams and evaluation against pre-treatment quality assurance results. *Phys Med Biol* 2015;60(6):2587–601.
- [48] McNiven AL, Sharpe MB, Purdie TG. A new metric for assessing IMRT modulation complexity and plan deliverability. *Med Phys* 2010;37(2):505–15.
- [49] Nicolini G, Clivio A, Cozzi L, Fogliata A, Vanetti E. On the impact of dose rate variation upon RapidArc® implementation of volumetric modulated arc therapy. *Med Phys* 2011;38(1):264–71.
- [50] Moustakis C, Blanck O, Ebrahimi Tazehmahalleh F, ka heng Chan M, Ernst I, Krieger T, et al. Planning benchmark study for SBRT of early stage NSCLC Planvergleichsstudie bei der SBRT des NSCLC im Frühstadium: ergebnisse der DEGRO AG Stereotaxie. *Strahlenther Onkol* 2017;193(10):780–90.
- [51] Buergy D, Sharfo AWM, Heijmen BJM, Voet PWJ, Breedveld S, Wenz F, et al. Fully automated treatment planning of spinal metastases – A comparison to manual planning of volumetric modulated arc therapy for conventionally fractionated irradiation. *Radiat Oncol* 2017;12(1). <https://doi.org/10.1186/s13014-017-0767-2>.
- [52] Wortel G, Eekhout D, Lamers E, van der Bel R, Kiers K, Wiersma T, et al. Characterization of automatic treatment planning approaches in radiotherapy. *Phys Imaging Radiat Oncol* 2021;19:60–5.