INTERNATIONAL JOURNAL OF
RADIATION ONCOLOGY • BIOLOGY • PHYSICS
www.redjournal.org

PHYSICS CONTRIBUTION

Multicenter End-to-End Evaluation of Single-Isocenter Multitarget Stereotactic Radiosurgery Approaches: Accuracy and Plan Quality

Benedikt Thomann, MSc,[a,b,c] Tobias Fechter, PhD,[a,b,c] Johannes Fischer, PhD,[b,d] Armin Runz,[e] Ute Ludwig, PhD,[b,d] Ilias Sachpazidis, PhD,[a,b,c] Oliver Blanck, PhD,[f] Julian Roers, PhD,[g] Melanie Grehn, MSc,[f] Maximilian Grohmann, MSc,[h] Christian Ziemann, MSc,[i] Michael Judge, MSc,[j] Wolfgang Baus, PhD,[k] Michelle Grahle, MSc,[l] Matthias Walke, PhD,[m] Bastian Bathen, PhD,[n] Janett Köhn, PhD,[n] Paul Käthner, MSc,[o] Maya Shariff, PhD,[p] Nils Wegner, MSc,[q] Jens Fleckenstein, PhD,[r] Heiko Karle, MSc,[s] Tino Streller, MSc,[t] Simon Howitz, MSc,[u] Marlen Priegnitz, MSc,[v] Rocco Weigel, MSc,[w] Thomas Konrad, MSc,[x] Daniela Schmitt, PhD,[y] Jürgen Beck, MD,[z] Marcia Machein, MD,[z] Ilinca Popp, MD,[b,c,aa] Michael Reiner, PhD,[ab] Christos Moustakis, PhD,[ac] Christian P. Karger, PhD,[d] Evangelos Pappas, PhD,[ad,ae] Anca-Ligia Grosu, MD,[b,c,aa] Michael Bock, PhD,[b,d] and Dimos Baltas, PhD[a,b,c]

[a]Division of Medical Physics, Department of Radiation Oncology, Medical Center - University of Freiburg, Freiburg, Germany; [b]Faculty of Medicine, University of Freiburg, Freiburg, Germany; [c]German Cancer Consortium (DKTK), Partner Site DKTK-Freiburg, Germany; [d]Division of Medical Physics, Department of Diagnostic and Interventional Radiology, University Medical Center Freiburg, Freiburg, Germany; [e]Medical Physics in Radiation Oncology, German Cancer Research Center (DKFZ), Heidelberg, National Center for Radiation Research in Oncology (NCRO), Heidelberg Institute for Radiation Oncology (HIRO), Heidelberg, Germany; [f]Department of Radiation Oncology, University Medical Center Schleswig-Holstein, Campus Kiel, Saphir Radiosurgery Center Frankfurt and Northern Germany, Kiel, Germany; [g]Department of Radiotherapy and Radiation Oncology, University Hospital Münster, Münster, Germany; [h]Department of Radiotherapy and Radiation Oncology, University Hospital Hamburg-Eppendorf, Hamburg, Germany; [i]Department of Radiotherapy, University Hospital Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; [j]Department of Radiation Oncology, Cyberknife and Radiation Therapy, Faculty of Medicine and University Hospital of Cologne, University of Cologne, Cologne, Germany; [k]Department of Radiotherapy, Wege Hospital Bonn, Bonn, Germany; [l]Department of Radiotherapy - Medical Physics, Städtisches Klinikum Dresden, Dresden, Germany; [m]University Clinic for Radiation Therapy, Medical Physics, Magdeburg, Germany; [n]Department of Radiation Oncology, University Hospital Frankfurt, Saphir Radiosurgery Center Frankfurt and Northern Germany, Frankfurt, Germany; [o]Department of Radiotherapy and Radiation Oncology, University Hospital Würzburg, Würzburg, Germany; [p]Department of Radiation Oncology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuernberg (FAU), Comprehensive Cancer Center Erlangen-EMN (CCC ER-EMN), Erlangen, Germany; [q]Department of Radiotherapy and Radiation Oncology - Medical Physics, Klinikum Stuttgart, Stuttgart, Germany; [r]Department of Radiotherapy and Radiation Oncology, University Medical Center Mannheim, Mannheim, Germany; [s]Department of Radiation Oncology and Radiotherapy, University Medical Center Mainz, Mainz, Germany; [t]Department of Radiation Oncology - Medical Physics, Lucerne Cantonal Hospital, Lucerne, Switzerland; [u]Department of Radiotherapy and Radiation Oncology, SRH Wald-Klinikum Gera, Gera, Germany; [v]Department of

Corresponding author; E-mail: benedikt.thomann@uniklinik-freiburg.de

Author Responsible for Statistical Analysis: Benedikt Thomann was responsible for the statistical analyses.

Disclosures: E.P. contributed scientifically to this study and is the chief scientific officer at RTsafe, whose products were used. All other authors declare no conflicts of interest. Financial support from German Cancer Aid (grant number 701143317) for this study is gratefully acknowledged. Whole-body magnetic resonance imaging systems (Deutsche Forschungsgemeinschaft DFG [INST 39/1365-1 FuGG], CoreFacility "Magnetic Resonance Development and Application Center Freiburg," Department of Radiology, University Medical Center Freiburg, Faculty of Medicine, University of Freiburg) were used to perform magnetic resonance imaging measurements.

Data Sharing Statement: Research data are stored in an institutional repository and will be shared on request to the corresponding author.

Acknowledgments—The authors thank German Cancer Aid for funding this project and the reviewers for their time, careful evaluation, and constructive feedback, which substantially improved the clarity and quality of the manuscript.

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ijrobp.2026.03.055.

Int J Radiation Oncol Biol Phys, Vol. 000, No. 00, pp. 1–13, 2026
0360-3016/$ - see front matter © 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
https://doi.org/10.1016/j.ijrobp.2026.03.055

Radiation Oncology and Radiotherapy, University Medical Center Lausitz - Carl Thiem, Cottbus, Germany; [w]University Medical Center for Radiotherapy and Radiation Oncology, Tirol Kliniken, Innsbruck, Austria; [x]Department of Therapeutic Radiology and Oncology, University Hospital Graz, Graz, Austria; [y]Department of Radiotherapy and Radiation Oncology, Comprehensive Cancer Center (G-CCC), University Medical Center Göttingen, Göttingen, Germany; [z]Department of Neurosurgery, Medical Center - University of Freiburg, Freiburg, Germany; [aa]Department of Radiation Oncology, Medical Center - University of Freiburg, Freiburg, Germany; [ab]Department of Radiation Oncology, University Hospital, LMU Munich, Munich, Germany; [ac]Department of Radiotherapy - Medical Physics, University Hospital Leipzig, Leipzig, Germany; [ad]Department of Biomedical Sciences, Radiology and Radiotherapy Sector, University of West Attica, Athens, Greece; and [ae]School of Biomedical Engineering & Imaging Sciences, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom

Received Oct 16, 2025; Revised Mar 6, 2026; Accepted for publication Mar 22, 2026

Purpose: We conducted a multicenter study on single-isocenter multitarget stereotactic radiosurgery to dosimetrically assess end-to-end test results and identify approaches and techniques influencing spatial accuracy and treatment plan quality.

Methods and Materials: An anthropomorphic head phantom with radiochromic film and polymer gel inserts was used with a reference structure set of 5 brain metastases. End-to-end tests were performed on-site at 23 centers in Germany, Austria, and Switzerland, each following its own single-isocenter multitarget stereotactic radiosurgery protocol. Spatial accuracy was quantified by comparing planned and measured prescription isodose-volume centroids. Plan quality was assessed from treatment planning system calculations using the Paddick gradient index (GI) and Paddick conformity index. Statistical analyses, including a generalized linear model, correlated results with protocol parameters to identify favorable systems and techniques.

Results: The mean spatial offset between measured and calculated prescription isodose centroids across all centers and targets was $0.9 \pm 0.4_{(1\sigma)}$ mm. Offsets above 1 mm were observed in 33% of centers. Better imaging-to-radiation isocenter consistency (ICC) yielded significantly higher accuracy ($P = .002$): $d_{ICC < median} = 0.6 \pm 0.2$ mm versus 1.1 ± 0.3 mm. Mean GI was 6.7 ± 3.3. Automated planning (AP) tools achieved significantly lower GI (4.9 ± 0.7) than conventional planning (8.1 ± 3.8, $P = .015$). Mean Paddick conformity index was 0.75 ± 0.17, with AP significantly improving conformity (0.83 ± 0.07 vs 0.68 ± 0.19; $P = .028$) and reducing variability in both indices. Target-to-isocenter distance had no significant influence on spatial accuracy, GI, or Paddick conformity index.

Conclusions: Spatial accuracy in a static phantom was primarily determined by ICC and less by specific delivery infrastructure or techniques, emphasizing the importance of a precise imaging isocenter calibration. AP tools significantly improved and standardized treatment plan quality across centers. © 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Introduction

Single-isocenter, multitarget (SIMT) stereotactic radiosurgery (SRS) has emerged as an efficient approach for simultaneously treating multiple brain metastases, with the potential to reduce treatment time while maintaining comparable plan quality and clinical outcomes.[1,2] Given these advantages, many clinics have already adopted SIMT SRS, whereas others are actively considering its implementation. However, the approach places stringent demands on treatment planning and dose delivery precision because even small geometric uncertainties can compromise target coverage and normal tissue sparing.[3]

Although guidelines exist for stereotactic radiation therapy and SRS in general,[4-7] specific recommendations for SIMT SRS are currently lacking. To address this gap, we conducted a multicenter, multiplatform study initiated by the working group for stereotactic radiation therapy of the German Society for Medical Physics (DGMP), in close collaboration with the working groups for neuroradiation oncology and for stereotactic radiation therapy of the German Society for Radiation Oncology, with funding from German Cancer Aid. The first phase of this study involved a systematic analysis of SIMT SRS infrastructure and protocol parameters across 23 radiation therapy centers in Germany, Austria, and Switzerland.[8] Building on these findings, the present and main part of the study evaluated the spatial accuracy of dose delivery and the clinical plan quality achievable with these systems and protocols. The aim was to identify potentially beneficial or detrimental approaches and factors that could inform future SIMT-SRS-specific guidelines.

Methods and Materials

Measurement phantom and reference case

An anthropomorphic head phantom (Prime, RTsafe) with inserts for radiochromic film and polymer gel dosimetry was used for end-to-end test measurements.[9,10] Except for the 3-dimensional (3D)-printed bone structures and the insert, the phantom was filled with water. The film insert consisted of a polystyrene casing holding a radiochromic film (7 × 14 cm, type EBT-3, Ashland Inc) placed between 2 polystyrene plates, enabling measurements in either the

coronal or sagittal plane. The gel insert consisted of a 3D-printed cylinder filled with polymer gel (8 cm in diameter and 15 cm in length).[11,12] In a reference computed tomography (CT) image, 5 spherical target structures (8 mm in diameter) were delineated at varying distances to the presumed isocenter (up to 5 cm to the center of mass of all targets). All target structures were fully encompassed by the gel volume and located on either the coronal or sagittal film plane, with the central target volume (TV) lying on both planes (see Fig. 1). Measurements with gel and the 2 film orientations were always performed consecutively.

Multicenter ring trial

A multicenter ring trial was conducted across 23 participating institutions in Germany, Austria, and Switzerland, including 3 CyberKnife (CK) centers (Accuray Inc). Among the 20 centers with C-arm linear accelerators, 10 had prior clinical experience with SIMT SRS (experienced [Exp] group), whereas the remaining 10 used the study to benchmark their preliminary protocols (benchmarking [Bench] group). A detailed analysis of the submitted protocols is available in a previous publication.[8]

Each center was visited with the phantom, serving as a pseudo-patient for a comprehensive end-to-end test. This included all steps of the center-specific SIMT SRS workflow: immobilization, imaging, treatment planning, and delivery. A CT scan was performed for each phantom configuration (1 gel insert and 2 film inserts). The 5 TVs were imported after co-registration with our reference CT data set. SIMT SRS treatment planning was performed on 1 CT data set, and the resulting plan was then recalculated on the remaining data sets. This approach ensured a correct setup reference for each configuration during pretreatment imaging, accounting for differences between inserts while still applying the same treatment plan in all measurements.

At each center, the workflow was completed in 1 day: imaging in the morning, followed by planning and irradiation in the afternoon or evening. Gel inserts were produced 2 days before irradiation and shipped on-site the day before irradiation. Films were prepared at the primary study site and transported on-site along with the phantom. Centers were visited in groups of 3 to 4 between Sunday and Wednesday. Film and gel readouts were performed the following Friday at the primary study site. All measurements were completed within approximately 4 months.

For treatment planning, the only requirement was a maximum dose (D_{max}) of ~10 Gy, corresponding to the dose range reliably covered by both film and gel calibration. As a result, the only modification to the center-specific SIMT SRS protocols enforced by our study was a simple dose rescaling, usually resulting in prescription dose levels of 7 to 8 Gy to encompass the TVs. All other planning parameters were determined by the participating centers according to their standard protocols. Anonymized data sets (CT, dose, plan, and structures per center) are deposited in an open-access online repository.[13] Infrastructure and protocol parameters are listed in Appendix E6 to enable external replication and further investigation.

Film dosimetry workflow

Film dosimetry followed the recommendations of the American Association of Physicists in Medicine,[14] using EBT-3 film (Ashland Inc) and the radiochromic.com platform.[15,16] Calibration was performed at the primary study site. To account for the 4-month duration of the ring trial, calibration was repeated regularly to ensure that no measurement used a calibration older than 4 weeks. Our detailed film calibration protocol is provided in Appendix E1.

The head phantom's film insert uses 4 metal pins to pierce and fixate the film between 2 polystyrene plates,

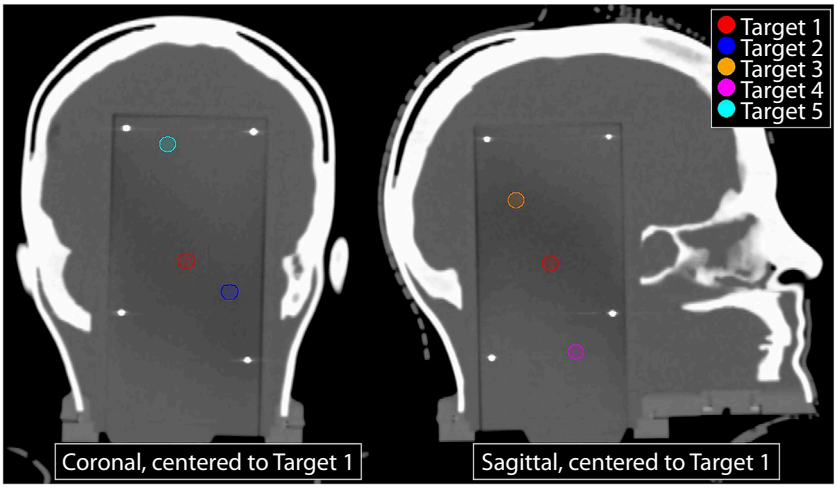


Fig. 1. The reference case shown in computed tomography images acquired with coronal and sagittal film insert orientations. Target structures were positioned such that the center-to-center distance from target 1 increases stepwise: ≈2 cm (target 2), ≈3 cm (target 3), ≈4 cm (target 4), and ≈5 cm (target 5).

which are then inserted into the phantom for irradiation. Films were scanned at 300 dpi on an Epson Expression 12000 XL scanner (Epson) and converted to Digital Imaging and Communications in Medicine (DICOM) dose planes using time-matched calibration data (see Appendix E1). In-house Python scripts for 3D Slicer[17] co-registered these dose planes with the corresponding treatment planning system (TPS) doses by aligning the 4 metal pins in the CT image with the corresponding holes in the film dose planes.

Gel dosimetry workflow

Polymer gel dosimetry[11] is based on dose-dependent changes in the relaxation time T_2, which were measured at the voxel level using magnetic resonance imaging (MRI) at the primary study site. These changes were translated into dose values through a dedicated calibration procedure, in which T_2 values were matched to known irradiation doses. All scans were performed using a 20-channel head coil on a clinical 3-T system (Cima.X, Siemens Healthineers), with the gel insert located inside the head phantom. To position the phantom inside the head coil, we used a 3D-printed head shell, ensuring the same setup in every measurement. Four multimodality CT/MRI markers (PinPoint, Beekley Medical) were permanently positioned inside the phantom, allowing co-registration between MRI and CT images. To identify the markers in the MR images, a 3D turbo spin echo sequence with modulated refocusing flip angles (SPACE, Siemens Healthineers) was used.[18,19] The dose-dependent T_2 changes were measured using a Half-Fourier-Acquired Single-shot Turbo spin Echo (HASTE) sequence at multiple echo times (echo time = 47, 55, 117, 203, 391, 712 ms, repetition time = 2800 ms, flip angle = 140°).[20] All MR data were acquired at 1 mm isotropic spatial resolution, and standard distortion correction was applied to account for gradient nonlinearities.

Shipped in a cooling box, every participating center received 2 gel inserts from the same production batch: one for dose calibration and one for SIMT SRS measurement. Performing calibration on-site with a second gel insert ensured identical temperature history and irradiation timing for both inserts. This inherently accounted for temperature variations during calibration, which might otherwise have altered gel chemistry and T_2 relaxation times. Our detailed gel calibration protocol is provided in Appendix E2.

The SIMT SRS insert followed the same workflow as the calibration insert but was irradiated with the clinical SIMT SRS plan instead of the calibration field. Image registration was performed as described in Appendix E2, and relaxation rates were converted to dose using the calibration curve derived from the corresponding calibration insert.

Dosimetric evaluations

All dosimetric evaluations were performed using in-house Python scripts and 3D Slicer. Evaluations were identical for film and gel, except for the inherent 2-dimensional (2D) limitation of film. In the following sections, "volume" refers to the corresponding plane in the case of film. Dose matrices were resampled to enable voxel-based comparison between measured and TPS-calculated dose. We present 3 main analyses: spatial dose delivery accuracy based on measured-to-planned dose comparisons, as well as gradient index (GI) and conformity index evaluations based on TPS calculations.

Spatial accuracy

The spatial accuracy of the delivered dose was the primary focus of this study. To quantify spatial offsets, the center-specific planning target volume (PTV)-encompassing prescription isodose regions were segmented in both the measured and TPS-calculated dose distributions, resulting in 5 segments per data set (1 per target). Offsets were calculated as the distances between the centroids of each corresponding segment pair and visualized in dose profiles as well as 2D dose difference maps ($D_{Film} - D_{TPS}$).

Although gel inherently provided 3D spatial offsets, the 3D offset d_{3D} for film was estimated as the root-sum square of offsets from the coronal and sagittal measurement planes, d_{cor} and d_{sag}. Because only the central TV was captured in both planes, the second component for the remaining 4 targets was estimated using the mean offset $\overline{d}_{TV_{All}}$ from all targets covered in the corresponding orthogonal plane. The calculation is shown in Equation 1 for a target covered only in the coronal plane and applies analogously for the sagittal plane. This approach is based on the observation that, within each center, offsets showed consistent magnitude and direction across all targets, indicating systematic rather than individual deviations (see Fig. 4B for an illustrative example).

$$d_{cor3D,TV_i} = \sqrt{\left(d_{cor,TV_i}\right)^2 + \left(\overline{d}_{sag,TV_{All}}\right)^2} \qquad (1)$$

Measurement uncertainty and reproducibility

The reliability of the results was assessed through repeated measurements at the primary study site, representing a type A evaluation of uncertainty based on statistical analysis of repeated observations.[21] For film dosimetry, the same SIMT SRS irradiation was performed 10 times in a single session, both for the coronal and the sagittal film insert orientation. The measured spatial offsets were used to calculate the SD across all 10 measurements, respectively. Because of the higher logistical complexity and cost of gel dosimetry, only 5 different SIMT SRS irradiations conducted as part of the study were repeated once during the same session. For each pair, the difference in spatial offsets was determined, and the mean and SD of these differences were calculated across all 5 measurement pairs. Type B uncertainties were estimated based on individual workflows.

Signed centroid offsets across all centers along the 3 anatomic axes were visualized with boxplots to show their

distribution and potential directional bias. One-sample t tests[22] against 0 were performed to assess whether mean offsets were statistically significant, with SDs and 95% CIs reported for each axis.

Agreement between film and gel measurements

To assess the agreement between film and gel dosimetry in determining spatial offsets between measured and calculated dose distributions, a Bland-Altman analysis was performed.[23] For each measurement point, the mean of the film and gel offsets $\overline{d}_{\text{Film\&Gel}}$ was plotted against their difference $\Delta_{d_{\text{Film}-\text{Gel}}}$. The mean difference (bias) and the 95% limits of agreement (approximated by $\pm 2\sigma$) were calculated to evaluate the interchangeability between the 2 methods. Additionally, the mean and SD σ of the spatial offset differences (film − gel) were calculated separately for each TV.

Treatment plan quality

In accordance with the recommendations of the International Commission on Radiation Units and Measurements,[24] Paddick GI and Paddick conformity index (PCI)[25,26] were calculated from TPS data of each center to assess treatment plan quality, using in-house Python scripts and 3D Slicer.

The GI was calculated from the TPS dose distribution, where prescription isodose volume (PIV) represents the volume receiving at least the prescription dose and PIV_{50} represents the volume receiving at least 50% of the prescription dose.[25]

$$\text{GI} = \frac{\text{PIV}_{50}}{\text{PIV}} \tag{2}$$

GI was calculated globally for the entire dose matrix and, where possible, target-specific basis for each target. Target-specific evaluation of the GI was not feasible in 7 cases in which the dose between adjacent targets did not fall below 50% of the prescription dose (1 Exp and 6 Bench centers).

The PCI was calculated from the TPS dose distribution for each target, where TV is the total TV and TVPIV is the TV covered by the prescription isodose.[26] Target-specific PCI calculation was not feasible in 1 case where only a union PTV was available because of optimization and reporting practices.

$$\text{PCI} = \frac{(\text{TV}_{\text{PIV}})^2}{\text{TV} \cdot \text{PIV}} \tag{3}$$

End-to-end test reporting

Each center received a report including the spatial offset for each TV, 2-dimensional dose difference maps in 2 planes ($\text{D}_{\text{Film}} - \text{D}_{\text{TPS}}$), GI and PCI, measured dose profiles compared with TPS-calculated profiles, and TPS-calculated profiles compared across centers. In the following sections, we focus on overall study results and selectively show individual results for illustrative purposes. A representative, full report for one of the participating centers is provided in Appendix E3.

Analysis of influencing factors

The influence of the target-to-isocenter distance on spatial offsets, GI, and PCI was investigated using 2-sided Student t tests.[22] Additionally, we assessed the influence of SIMT SRS infrastructure and protocol parameters, details of which have been published previously.[8] These include linear accelerator and collimator models, couch models, TPSs, planning techniques and parameters, intrafractional imaging protocols, and quality assurance (QA) procedures. All parameters were considered as binary variables, indicating whether a specific technique or approach was used. Appendix E6 summarizes the key parameters and their implementation across participating clinics. To assess the potential influence of these parameters on the results, appropriate statistical tests were chosen based on data distribution and variance characteristics. Specifically, normality and variance equality were assessed for each parameter group, guiding the selection between parametric tests (2-sample t test or Welch t test[27]) when assumptions were met and the nonparametric Mann-Whitney U test otherwise.[28] These tests compared the distributions of results (mean spatial offset, mean GI, and mean PCI of each center) between groups defined by the presence or absence of a given technique. We report corresponding Pvalues and Cohen d effect sizes,[29] considering results statistically significant at $\alpha = 0.05$.

To support and contextualize these findings, generalized linear models (GLM) were applied as secondary analyses. Because of the limited sample size (23 centers), the GLM served mainly to indicate trends rather than provide definitive statistical results. Each GLM initially included the parameter of interest, a second potentially relevant parameter that showed significance in the t test (eg, the center's experience status), and their interaction term. This was done to identify whether any observed effect could be confounded by other parameters or whether it persisted independently. GLM residuals were assessed for normality to evaluate the validity of the model assumptions. All analyses were performed in Python using the numpy (v1.25.1), scipy (v1.11.1) and statsmodels (v0.14.4) libraries.[30-32] Scripts will be made available on request.

Results

Measurement uncertainties

Overall uncertainties for film and gel were derived from type A (repeatability) and type B (systematic) components.[21] The dominant type B source was registration to the planned dose (CT). Although a baseline uncertainty of 0.5 mm (half the voxel size of CT/MRI) could be assumed, actual values were lower because of 4-point marker-based registration. Mean

root-mean-square registration uncertainties across 23 data sets were $u_{Reg,Film}$ = 0.25 mm and $u_{Reg,Gel}$ = 0.32 mm. For film, assuming uniform offsets across all targets introduced an additional type B uncertainty of $u_{uniform}$ = 0.13 mm, based on the root-mean-square of 23 SDs. Type A uncertainties were $u_{A,Film}$ = 0.15 mm (0.17 mm coronal, 0.14 mm sagittal) and $u_{A,Gel}$ = 0.18 mm. Overall uncertainties were calculated via root-sum-of-squares of the respective components and are reported as expanded uncertainties U at k = 2: U_{Film} = 0.64 mm and U_{Gel} = 0.73 mm.

Signed spatial offsets across centers, separated into their 3 spatial components (left−right, inferior−superior, and anterior−posterior) for the purpose of investigating potential directional bias, are summarized in Appendix E4.1. Film dosimetry showed no detectable directional bias along any axis. Gel measurements exhibited a small anterior−posterior offset of 0.2 mm that was statistically significant but minor compared with the overall variation in offsets and the measurement uncertainty. These results indicate that the measurements are reliable and not meaningfully biased.

Measurement agreement and reporting

The Bland-Altman analysis (Fig. 2) showed a mean bias of $\overline{\Delta}$ = −0.04 mm and limits of agreement of $2\sigma_{BA}$ = 0.82 mm, below the combined uncertainty of film and gel ($U_{Film\&Gel}$ = 0.97 mm), suggesting a possible overestimation of type B components. Differences between film and gel showed no dependency on magnitude or TV. The low bias indicates interchangeability of film and gel measurements for overall results.

Unfortunately, technical issues—likely because of oxygen contamination from in-house 3D-printed cylinders—reduced gel sensitivity. Although primarily affecting low-dose regions, the impact was most pronounced for target 4, located closest to the fill port, and prevented evaluation of 13% of all targets across centers. Consequently, the following sections focus on film results, and gel results are provided in the Appendix D5. Notably, all major trends and conclusions observed in the film data were consistently reflected and confirmed by the gel data, supporting the interchangeability of the 2 dosimeters.

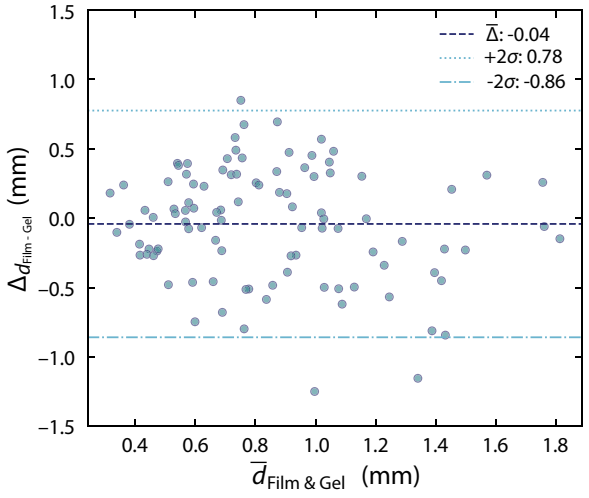


Fig. 2. Film and gel agreement: Bland-Altman analysis comparing spatial offsets from film and gel. $\Delta_{d_{Film\ -\ Gel}}$ is the spatial offset difference, and $\overline{d}_{Film\&Gel}$ is the mean spatial offset per target per center. The black dashed line ($\overline{\Delta}$) indicates the mean difference across all targets and centers.

Spatial accuracy

The mean overall spatial offset between measured and planned dose across all targets and centers was $\overline{d}_{C\text{-}arm}$ = 0.89 ± $0.41_{(1\sigma)}$ mm for the C-arm group and $\overline{d}_{CK}$ = 0.70 ± 0.29 mm for the CK group. Subgroup results are presented in the following sections and in Appendix E4. In this study, the applied PTV margins varied and did not always correspond to standard clinical practice. Some centers deliberately adjusted their margins, either reducing them to investigate the feasibility of smaller targets or increasing them when the resulting targets did not meet their minimum size requirements. The most common clinical PTV margin was 1 mm. Overall, 33% of measured target offsets exceeded 1 mm, and 7% exceeded 1.64 mm (PTV margin + U_{Film} at k = 2). Offsets >1 mm were observed in 11 of 23 centers, and 2 centers treated at least 1 target with an offset >1.64 mm.

Influence of infrastructure parameters

Within the C-arm group, 2-sample *t* tests identified 3 parameters significantly associated with improved accuracy: experience, the use of intrafraction imaging (IF-Img), and the consistency between imaging and treatment isocenters (see Fig. 3 for graphical representation with *P*values and Appendix E4.2 for numerical values). For the latter, the C-arm centers reported their imaging-to-radiation isocenter consistency (ICC) based on the QA sessions closest to the measurement date. Centers were divided into groups $ICC_{<Med}$ and $ICC_{>Med}$, depending on whether their ICC was below or above the median of ICC_{Med} = 0.24 mm. Four centers reported insufficient data and were excluded from this statistical evaluation.

A GLM analysis showed that the effects of experience and IF-Img were confounded by ICC (P_{Exp} = .240 vs P_{ICC} = .017 and $P_{IF\text{-}Img}$ = .726 vs P_{ICC} = .038). ICC alone remained independently significant. Further results of the GLM investigating these 3 parameters and their interaction are provided in Appendix E4.3. These findings suggest that overall spatial treatment accuracy was primarily driven by ICC in our study. Parameters not reported here were not found to have a significant influence on delivery accuracy (eg, the number of couch angles, couch system, or treatment unit).

Influence of target-to-isocenter distance

The distance between target and radiation isocenter had no significant impact on spatial accuracy, either overall or

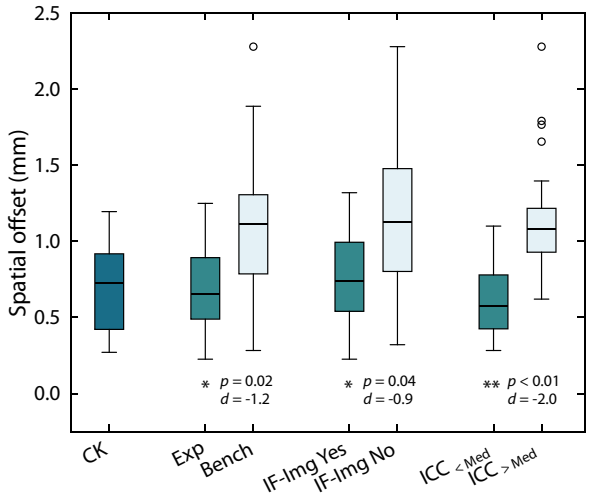


Fig. 3. Spatial offsets for all targets, analyzed for subgroups of centers: CyberKnife (CK) and C-arm centers: experienced (Exp), benchmarking (Bench), use of intrafraction imaging (IF-Img), and imaging-to-radiation isocenter consistency (ICC) level (below or above the median reported value). Boxes represent the IQR, the median (line), data range within 1.5 × IQR (whiskers), and outliers beyond whiskers (points). Significance level is shown by *P* values, effect size magnitude by Cohen d. Asterisks indicate significance in univariate (*) or multivariate (**) tests.

within defined subgroups, for both film and gel dosimetry. Figure 4A illustrates the relationship between measured spatial offset and target-to-isocenter distance, stratified by the centers' IF-Img use. Corresponding analyses for other subgroups (experience status and ICC) are provided in Appendix E4.4. The supplementary gel dosimetry analysis (Appendix E4.5) confirms these results. In addition, Figure 4B includes a dose difference map from 1 representative center, demonstrating consistent offsets in both magnitude and direction across all 5 targets. This pattern was observed in most centers and supports the conclusion that spatial treatment accuracy was less dependent on target-to-isocenter distance and more on systematic factors, such as ICC.

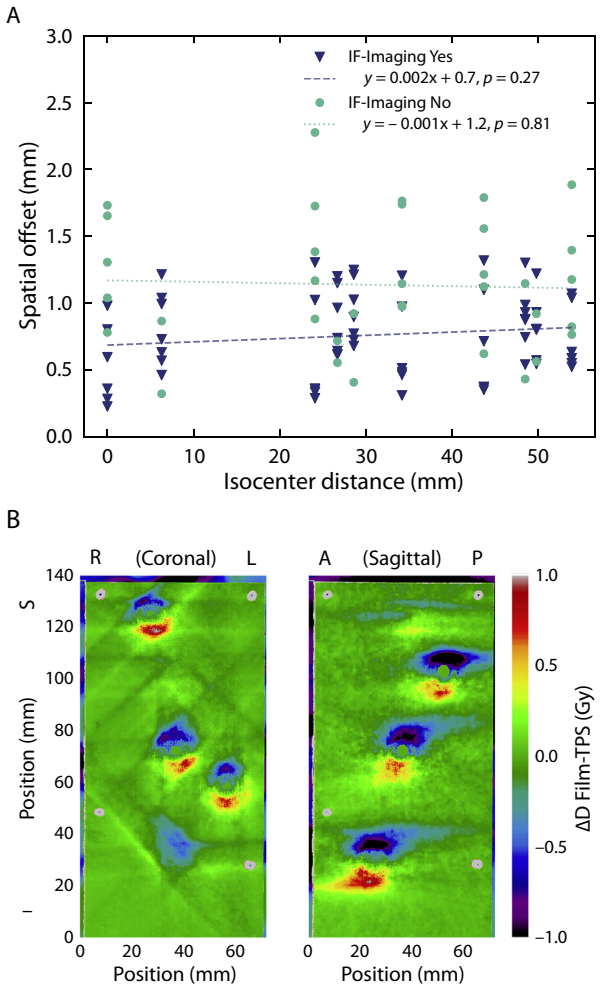


Fig. 4. (A) Measured spatial offsets of all targets across C-arm centers as a function of target-to-isocenter distance, stratified by the centers' use of intrafraction imaging (IF-Img). Regression lines were fitted per group; *P* values indicate no slope significance. (B) 2-Dimensional dose difference between film measurement and TPS calculation in the central coronal and sagittal planes for one representative center. The PTV-encompassing prescription dose was 8 Gy. Offsets are similar in magnitude and direction for all targets, suggesting independence from target-to-isocenter distance. *Abbreviations:* I = Inferior; S = Superior; R = Right; L = Left; A = Anterior; P = Posterior; TPS = treatment planning system.

Treatment plan quality

Qualitative overview

Figure 5 shows coronal film images (TV1, 2, and 5), as well as TPS profiles for TV1 from all centers. Because PTV margins and dose prescription protocols varied between centers, profile widths and the relation between prescription and D_{max} values inherently differ. Thus, Figure 5 should be interpreted qualitatively and with caution. Nevertheless, it highlights substantial variation in dose distributions across centers, particularly among C-arm centers without automated planning (AP) tools (HyperArc, Varian Medical Systems, a Siemens Healthineers company; and Elements, Brainlab SE).

Gradient index and conformity index

Figure 6 illustrates the GI and PCI results for relevant subgroups of centers, defined by experience status, planned D_{max}, multileaf collimator (MLC) system, and use of AP tools, as well as corresponding *P*values. Within the C-arm group, 3 parameters showed significantly improved results: AP, a smaller MLC leaf width of 2.5 mm (HD120 MLC, Varian Medical Systems), and a D_{max} above the median

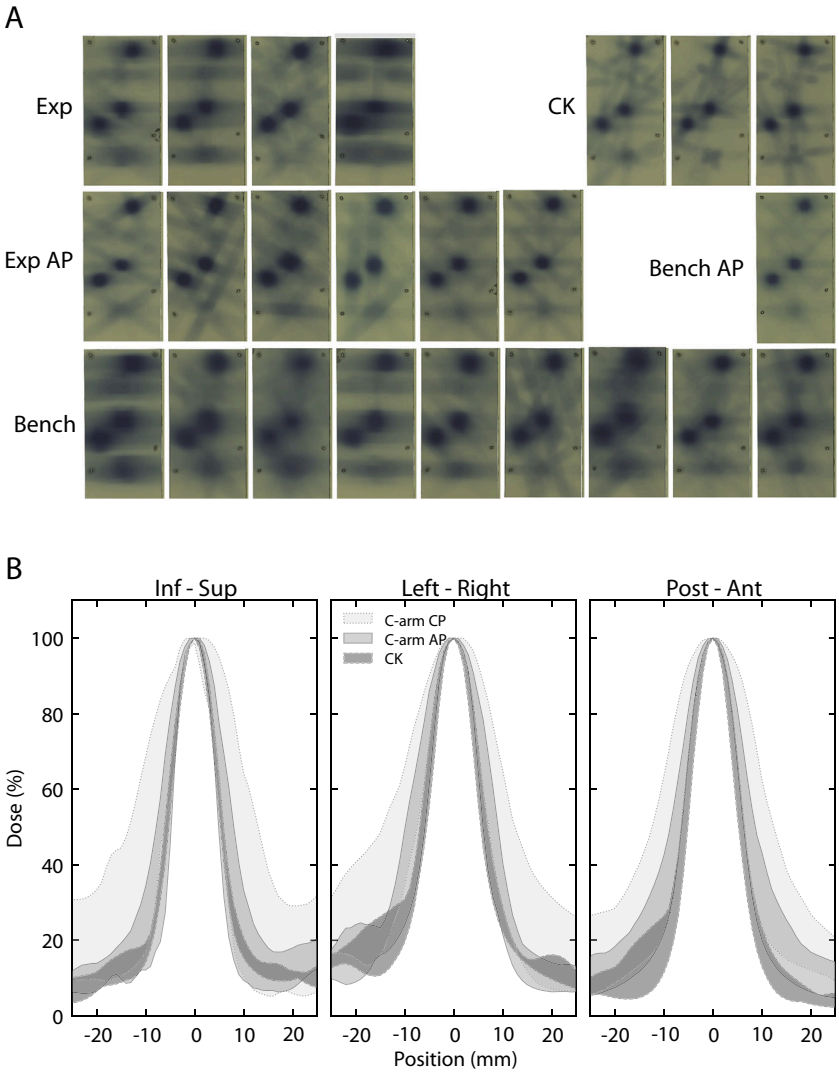


Fig. 5. Qualitative comparison of dose distributions across centers. C-arm centers were stratified by experience (experienced [Exp] vs benchmarking [Bench]) and planning technique (conventional planning (CP) vs. automated planning (AP)) and compared with CyberKnife (CK). PTV margins varied between centers. (A) Coronal film scans. (B) Treatment planning system profiles for the central target volume, normalized to 100% at the dose maximum and aligned with the individual prescription isodose centroid. *Abbreviations:* Ant = Anterior; Inf =Inferior; Sup =Superior.

D_{max} of all C-arm centers. Centers using AP tools achieved significantly lower GI (4.9 ± 0.7 vs 8.1 ± 3.8; P = .02) and significantly higher PCI (0.83 ± 0.07 vs 0.68 ± 0.19; P = .03) than centers using conventional planning. Centers using a 2.5-mm MLC system also achieved significantly lower GI values (4.3 ± 0.9 vs 7.9 ± 3.5; $P < .01$) and improved conformity (0.81 ± 0.10 vs 0.71 ± 0.19; P = .29). Centers with a higher planned D_{max} achieved significantly lower GI (5.8 ± 2.5 vs 8.2 ± 4.0; P = .05). More detailed numerical data for all subgroups are provided in Appendix E5.1.

In a multivariate GLM, the apparent effect of D_{max} was no longer significant after adjustment for AP, indicating that the observed association with GI was largely explained by the use of AP. Corresponding statistical results are provided in Appendix E5.3. Notably, AP was not associated with systematically higher D_{max} values (mean D_{max} 133% for AP vs 132% for non-AP centers, with AP centers evenly distributed across both median-based D_{max} groups), suggesting that the improved GI observed with AP was not achieved by simply increasing D_{max}.

Because of substantial group overlap (7 centers used AP, 5 used the HD120 MLC, and 4 used both) and nonnormally distributed residuals in the GLM, it was not possible to statistically separate individual contributions within or between AP and MLC design. Their individual partial effects are visualized in Appendix E5.2.

Comparing the 2 AP tools, HyperArc centers achieved higher conformity, whereas Elements centers achieved better gradients. Because of small samples and modest effect sizes, these results should be interpreted only as indicative trends.

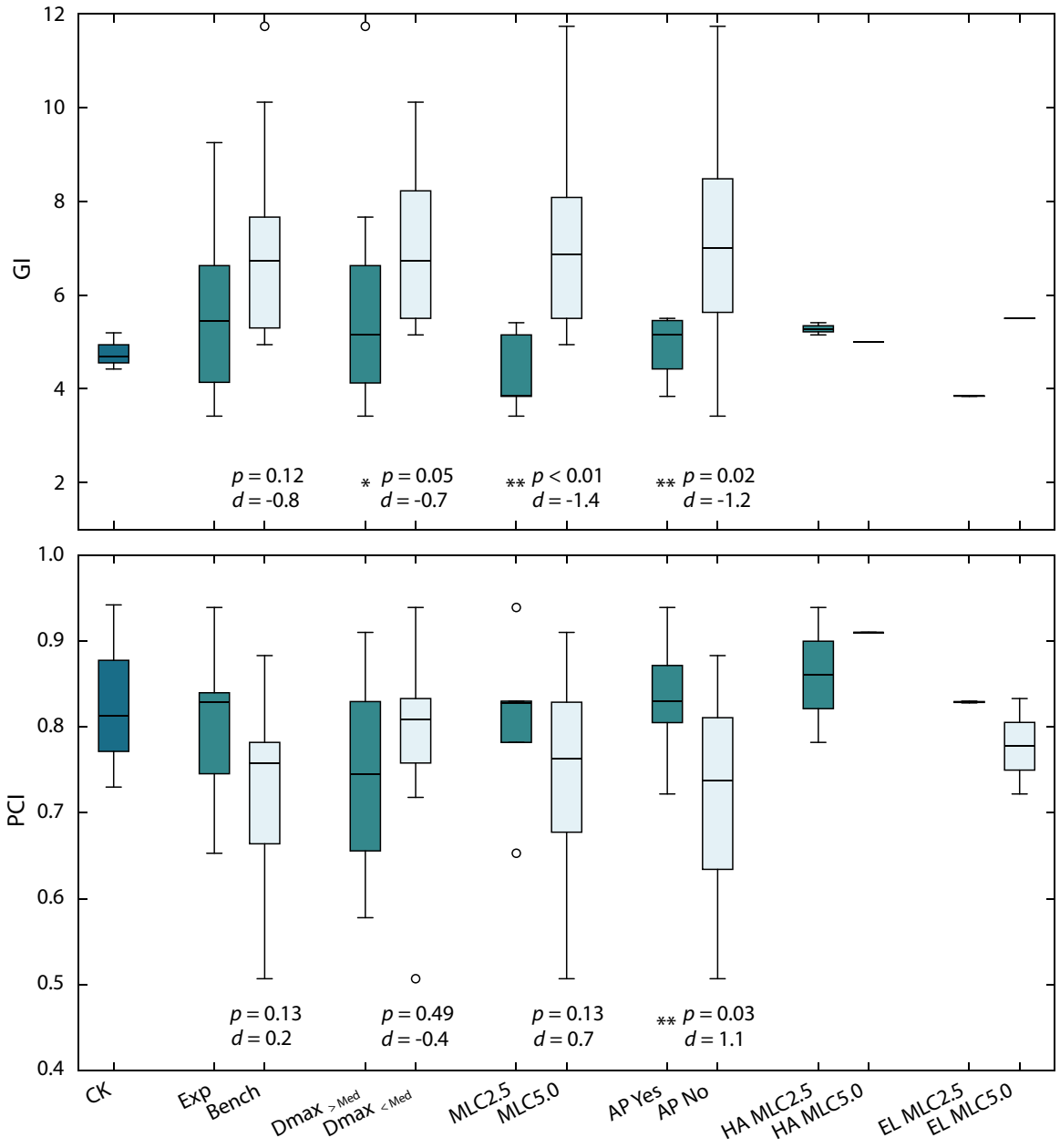


Fig. 6. Gradient index (GI) and Paddick conformity index (PCI) with statistical results for CyberKnife (CK) centers and C-arm center subgroups: experienced (Exp), benchmarking (Bench), maximum dose (D_{max}) (above or below the median value of C-arm centers), multileaf collimator (MLC) leaf width in mm (MLC2.5 and MLC5.0), and automated-planning (AP) systems (HyperArc [HA], Elements [EL]). Boxes represent the IQR, the median (line), data range within 1.5 × IQR (whiskers), and outliers beyond whiskers (points). Significance level is indicated by *P* values, effect size magnitude by Cohen d. Asterisks indicate significance in univariate (*) or multivariate (**) tests.

The lowest GI_{lowest} = 3.4 in our study was achieved by C-arm with a RayStation TPS (RaySearch Laboratories AB) and the HD120 MLC. The 2 highest $PCI_{highest}$ = 0.94 were achieved by a CK and a C-arm system with HyperArc AP and the HD120 MLC. Figure 6 illustrates that although comparable results were achieved within all subgroups with different infrastructure and techniques, the use of AP tools, in particular, enabled consistent, high-quality GI and PCI results with reduced user dependence. Target-to-isocenter distance, individual PTV margins, and all other parameters not reported here (see Appendix E6 for a complete list) were not found to have a significant influence on GI or PCI. An additional GLM including 3 PTV margin subgroups (<1, 1, and >1 mm) was performed to assess a potential confounding effect. No significant association between PTV margin and GI or PCI was observed (see Appendixes E5.4 and E6).

Dose bridging at the 50% isodose level between PTV1 and PTV2 (see Fig. 1 for anatomic reference) was observed in 7 centers, 6 of which were Bench centers, and none of

which were centers using AP. Notably, 3 of these centers used a PTV margin of 2 mm, resulting in a surface-to-surface distance of only 10 mm between the 2 TVs.

Discussion

Measurement system

Enabled by the anthropomorphic phantom,[9] both film and gel systems performed robustly in our end-to-end tests. Although results were largely interchangeable, using both systems provided redundancy and increased reliability. By aligning all targets with film planes in our reference case, we mitigated the main limitation of film—its planar nature—whereas gel inherently allows true 3D dosimetry for arbitrary target topographies.

The film 3D offset, derived via the root-sum-square of coronal and sagittal offsets, likely represents a conservative estimate, because it combines two 2D measures. A coordinate-wise calculation was not pursued because each film data set relied on its own CT scan with independent positioning and dose calculation, which would have introduced additional uncertainty. Because the estimated film results agreed closely with the gel results (Fig. 2), any potential overestimation from this approach is likely negligible.

Although gel measurements exhibited a small anterior −posterior bias (0.2 mm), this effect was minor relative to the overall variation and measurement uncertainty and did not affect the overall agreement between detector systems or the resulting conclusions.

Spatial accuracy

The absence of a significant influence of IF-Img on delivery accuracy for our static phantom suggests that the hardware components involved in treatment delivery, particularly rotating elements such as the collimator and couch, performed reliably in our study. However, because the phantom did not simulate intrafraction patient motion within the thermoplastic masks, these results should not be interpreted to mean that IF-Img is unnecessary. Previous studies have demonstrated the impact of intrafraction motion on SRS delivery accuracy[33] and recommend the use of IF-Img for SRS.[6,7] Within the 5 cm range examined in this study, our results show that delivery accuracy does not systematically decrease with increasing distance from the isocenter. This finding suggests that distance-dependent PTV margins may be unnecessary. However, it remains essential to assess both isocentric and eccentric accuracy when implementing SIMT SRS, to confirm these results for each system and make adjustments as needed.

These observations are consistent with the finding that ICC was the only unconfounded, statistically significant parameter influencing delivery accuracy in our study. Of the 23 centers, 14 reported verifying ICC daily when stereotactic radiation therapy was scheduled for that day, in line with recommendations from a DGMP and German Society for Radiation Oncology consensus statement.[6] However, the tolerance ranges applied for this QA, and consequently the achieved ICC, varied among centers, with a reported median of ICC_{Med} = 0.24 mm. Our findings support the necessity of this QA and further indicate that tightening ICC tolerances can directly improve treatment delivery accuracy.

Treatment plan quality

The significant variation in GI and PCI across centers highlights the complexity of SIMT SRS treatment planning. Although high-quality plan was achieved by individual centers in all subgroups, both AP tools and the HD120 MLC system (leaf width of 2.5 mm) had a significant positive influence on GI and PCI in our study. Benefits of the HD120 MLC for SRS have also been reported previously,[34−36] although not to the extent observed here.

Our results support prior reports that a major advantage of AP systems is their ability to produce consistently high-quality plans with reduced user dependence, enabling greater standardization both across institutions and within teams.[37] Although comparison of the 2 AP systems—HyperArc (3 centers) and Elements (4 centers)—was limited by sample size, the observed trends of Elements showing slightly better GI and HyperArc slightly better PCI are consistent with previous work.[37,38]

Limitations

Data analysis

GI and PCI were derived from centrally analyzed DICOM data sets provided by the participating centers. Although absolute values may differ from center-reported metrics because of sampling and algorithmic differences between 3D Slicer and local TPSs, this approach ensured consistent and standardized metric calculation across centers.

Intrafractional patient motion

Although all system-related translational and rotational uncertainties (gantry, couch, collimator, and imaging−registration chain) were directly measured, intrafractional patient motion was not included. Such motion could not be reproducibly simulated using the modular anthropomorphic head phantom within stereotactic mask systems.

Clinical outcome

The study did not include patient data or clinical outcomes. The potential clinical impact of observed variations in plan quality and delivery accuracy must therefore be inferred by comparison with published data. Clinical implications are discussed in the subsequent subsection.

TV and location

The experimental setup permitted a maximum intertarget distance of ~9 cm. Target diameters ranged from 8 to 12 mm, depending on PTV margin. Because neither PTV margin nor target-to-isocenter distance significantly influenced GI, PCI, or spatial offsets, and because offsets were systematic rather than target-specific, translation of our findings to other target topographies may be cautiously considered but remains constrained by the specific geometry and conditions of the present study.

Platform representation and sample size

This study included 23 centers (20 conventional C-arm linacs and 3 CK centers), and statistical findings should therefore be interpreted with caution. The limited number of participating centers reflects the substantial logistical and financial effort required for on-site end-to-end testing. Consequently, the results are restricted to the investigated platforms and do not directly extend to other dedicated SRS systems. Within these constraints, the data suggest comparable performance of Exp C-arm and CK centers when key standards such as high ICC and AP are applied.

Cost to deliver

Although auto-planning and hypofractionation have been associated with improved workflow efficiency and reduced treatment delivery time,[39] cost to deliver was not explicitly assessed or quantified in this study. Although our findings further support the technical benefits of auto-planning with respect to plan quality and standardization, they also underscore the importance of comprehensive and potentially more resource-intensive QA, as well as advanced planning infrastructure. These aspects should be considered in any comprehensive cost to deliver evaluation.

Protocol heterogeneity

Participating centers were intentionally allowed to apply their in-house SIMT SRS workflows without study-imposed constraints (eg, D_{max}, PTV margin, and beam configuration) to reflect current clinical practice. Although a more homogeneous protocol would have simplified interpretation and increased methodological control, this design enabled evaluation of plan quality and delivery accuracy as achieved under real-world conditions. Consequently, the results should be interpreted in the context of this intentional heterogeneity.

Clinical implications

The clinical implications of this study are 2-fold. First, the observed variations in spatial accuracy, GI, and PCI may have direct relevance for clinical care, although their impact must be inferred indirectly through comparison with published data. Second, the audit nature of this multi-institutional study provides a realistic snapshot of current SIMT SRS implementations, highlighting both the benefits of specific techniques and potential dosimetric warning signs, thereby underscoring the need for detailed guidelines and further standardization.

Spatial offsets above 1 mm in delivered dose were observed in 33% of centers (excluding measurement uncertainty), and dose bridging of the 50% isodose between PTV1 and PTV2 occurred in 30% of centers. Although dose bridging was partially attributable to PTV margins and target proximity, both targeting inaccuracies and unintended high-dose exposure of healthy tissue are of particular concern in single-fraction SRS, where such effects have been associated with increased toxicity and risk of radionecrosis.[40]

Collectively, these findings emphasize the technical complexity of SIMT SRS and support the use of comprehensive end-to-end testing and Bench against Exp centers or published reference data during clinical implementation.

Context and outlook

The principle of auditing is increasingly being adopted worldwide for new, complex techniques to ensure treatment safety and quality.[38,41–46] To the best of our knowledge, this is the first multiplatform SIMT SRS ring trial to incorporate end-to-end tests using 3D dosimetry for spatial accuracy analysis alongside TPS evaluations. The trial not only provided participating centers with the opportunity to benchmark their approaches and compare results with peers, but also offered a realistic representation of current clinical practice in SIMT SRS. The observed heterogeneity in results highlights the value of multicenter end-to-end testing and underscores the importance of such initiatives. To further promote transparency and reproducibility, anonymized planning data sets have been made available in an open-access online repository,[13] and an extensive Appendix is provided alongside this manuscript. Together, these resources allow external reproduction of the results, facilitate additional analyses, and support expansion of the data set in future studies.

Conclusions

In a multicenter, multiplatform study including 23 centers, SIMT SRS treatment delivery accuracy in a static head phantom significantly correlated with center experience, IF-Img, and imaging-to-radiation ICC, with the latter being the dominant factor in multivariate analyses. Insufficient imaging-to-radiation isocenter calibration was associated with spatial offsets exceeding 1 mm, which were observed in 33% of centers, highlighting the need for standardized and specific guidelines (eg, tolerance values for ICC), end-to-end testing, and audits. Spatial accuracy was independent of target-to-isocenter distance within the investigated range of 5 cm. Treatment plan quality was enhanced by auto-planning tools and a smaller MLC leaf width, resulting in improved and standardized conformity and dose gradients, as well as reduced high-dose bridging between closely spaced targets.

ARTICLE IN PRESS

References

1. Clark GM, Popple RA, Young PE, Fiveash JB. Feasibility of single-isocenter volumetric modulated arc radiosurgery for treatment of multiple brain metastases. *Int J Radiat Oncol Biol Phys.* 2013;86:426-432.
2. Palmer JD, Sebastian NT, Chu J, et al. Single-isocenter multitarget stereotactic radiosurgery is safe and effective in the treatment of multiple brain metastases. *Adv Radiat Oncol.* 2019;5:70-76.
3. de Almeida Ribeiro C, Breschi L, de Donato A, et al. Dosimetric impact of setup errors on stereotactic technique with hyperArc for multiple brain lesions using single-isocenter and non coplanar arcs. *Eur J Med Phys* 2024;125S1:103652.
4. Halvorsen PH, Cirino E, Das IJ, et al. AAPM-RSS medical physics practice guideline 9.a. for SRS-SBRT. *J Appl Clin Med Phys.* 2017;18:10-21.
5. Cirino E, Benedict SH, Dupre PJ, et al. AAPM-RSS medical physics practice guideline 9.b: SRS-SBRT. *J Appl Clin Med Phys.* 2025;26:e14624.
6. Guckenberger M, Baus WW, Blanck O, et al. Definition and quality requirements for stereotactic radiotherapy: consensus statement from the DEGRO/DGMP Working Group Stereotactic Radiotherapy and Radiosurgery. *Strahlenther Onkol* 2020;196:417-420.
7. Schmitt D, Blanck O, Gauer T, et al. Technological quality requirements for stereotactic radiotherapy. *Strahlenther Onkol* 2020;196:421-443.
8. Thomann B, Fechter T, Fischer J, et al. Multicenter multiplatform pattern-of-practice analysis of single-isocenter multitarget stereotactic radiosurgery. *Strahlenther Onkol* 2025;201(9):953-962. https://doi.org/10.1007/s00066-025-02424-w.
9. Makris DN, Pappas EP, Zoros E, et al. Characterization of a novel 3D printed patient specific phantom for quality assurance in cranial stereotactic radiosurgery applications. *Phys Med Biol* 2019;64:105009.
10. Pappas E, Makris D, Zoros E, et al. Personalized end-to-end QA in cranial SRS: evaluation of the phantom-to-patient dosimetric equivalency of a 3D printed phantom using film dosimetry. *Int J Radiat Oncol Biol Phys* 2018;102(Suppl):e485-e486.
11. Pappas E, Maris T, Angelopoulos A, et al. A new polymer gel for magnetic resonance imaging (MRI) radiation dosimetry. *Phys Med Biol* 1999;44:2677.
12. Pappas E, Angelopoulos A, Kipouros P, Vlachos L, Xenofos S, Seimenis I. Evaluation of the performance of VIPAR polymer gels using a variety of x-ray and electron beams. *Phys Med Biol* 2003;48:N65.
13. Thomann B. Online TPS data repository for 23 participating centers; *Zenodo EU Open Research Repository*. In: *https://doi.org/10.5281/zenodo.18257314*.
14. Niroomand-Rad A, Chiu-Tsao S-T, Grams MP, et al. Report of AAPM task group 235 radiochromic film dosimetry: anupdate to TG-55. *Med Phys* 2020;47:5986-6025.
15. Radiochromic.com— Web *Application for Film Dosimetry and QA*. https://www.radiochromic.com, Accessed March 6, 2026.
16. Méndez I, Polšak A, Hudej R, Casar B. The Multigaussian method: a new approach to mitigating spatial heterogeneities´ with multichannel radiochromic film dosimetry. *Phys Med Biol* 2018;63:175013.
17. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imag* 2012;30:1323-1341.
18. Busse RF, Hariharan H, Vu A, Brittain JH. Fast spin echo sequences with very long echo trains: design of variable refocusing flip angle schedules and generation of clinical T2 contrast. *Magn Reson Med* 2006;55:1030-1037.
19. Busse RF, Brau AC, Vu A, et al. Effects of refocusing flip angle modulation and view ordering in 3D fast spin echo. *Magn Reson Med* 2008;60:640-649.
20. Semelka RC, Kelekis NL, Thomasson D, Brown MA, Laub GA. HASTE MR imaging: description of technique and preliminary results in the abdomen. *J Magn Reson Imag* 1996;6:698-699.
21. Joint Committee for Guides in Metrology (JCGM). *Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement*. JCGM; 2008.
22. Gosset WS. The probable error of a mean. *Biometrika* 1908;6:1-25.
23. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986;327:307-310.
24. Wilke L, Andratschke N, Blanck O, et al. ICRU report 91 on prescribing, recording, and reporting of stereotactic treatments with small photon beams: statement from the DEGRO/DGMP working group stereotactic radiotherapy and radiosurgery. *Strahlenther Onkol* 2019;195:193-198.
25. Paddick I, Lippitz B. A simple dose gradient measurement tool to complement the conformity index. *J Neurosurg* 2006;105(Suppl):194-201.
26. Paddick I. A simple scoring ratio to index the conformity of radiosurgical treatment plans. *J Neurosurg* 2000;93(Suppl 3):219-222.
27. Welch BL. The generalization of 'student's' problem when several different population variances are involved. *Biometrika* 1947;34:28-35.
28. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 1947;1850-1860.
29. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates; 1988.
30. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature* 2020;585:357-362.
31. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods* 2020;17:261-272.
32. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. *Proc 9th Python Sci Conf* 2010;57:61.
33. Grohmann M, Todorovic M, Petersen C. Der Einfluss von intrafraktioneller Patientenbewegung auf die Behandlungsgenauigkeit bei Radiochirurgie multipler Hirnmetastasen. *Abstractband der DEGRO-Jahrestagung* 2021;197(Suppl 1):S64.
34. Chae SM, Lee KW, Son SH. Dosimetric impact of multileaf collimator leaf width according to sophisticated grade of technique in the IMRT and VMAT planning for pituitary adenoma lesion. *Oncotarget* 2016;7:78119-78126.
35. Chae SM, Lee KW, Son SH. The effect of multileaf collimator leaf width on the radiosurgery planning for spine lesion treatment in terms of the modulated techniques and target complexity. *Radiat Oncol* 2014;9:72.
36. Jalbout W, Abou Zahr J, Youssef B, Shahine B. On the feasibility of stereotactic radiosurgery with 5.0 and 10.0 mm MLC leaves as a function of target size and shape. *Front Oncol* 2019;9:741.
37. Todorovic M, Grohmann M, Frenzel T, Petersen C. *Dosimetrische Evaluation zweier moderner automatischer Bestrahlungsplanungslösungen für die Radiochirurgie multipler Hirnmetastasen* 2021;197:S140.
38. Paddick I, Mott J, Bedford J, et al. Benchmarking tests of contemporary SRS platforms: have technological developments resulted in improved treatment plan quality? *Pract Radiat Oncol* 2023;13:e451-e459.
39. Popple RA, Brown MH, Thomas EM, et al. Transition from manual to automated planning and delivery of volumetric modulated arc therapy stereotactic radiosurgery: clinical, dosimetric, and quality assurance results. *Pract Radiat Oncol* 2020;11:e163-e171.
40. Milano MT, Grimm J, Niemierko A, et al. Single- and multifraction stereotactic radiosurgery dose/volume tolerances of the brain. *Int J Radiat Oncol Biol Phys.* 2021;110:68-86.
41. Eaton DJ, Lee J, Paddick I. Stereotactic radiosurgery for multiple brain metastases: Results of multicenter benchmark planning studies. *Prac Radiat Oncol* 2018;8:e212-e220.
42. Dimitriadis A, Tsang Y, Thomas RAS, et al. Multi-institutional dosimetric delivery assessment of intracranial stereotactic radiosurgery on different treatment platforms. *Radiother Oncol* 2020;147:153-161.
43. Martucci P, Embriaco A, Pimpinella M, et al. Dosimetry audit service in Italy: results of the partnership between Italian Association of Medical Physics (AIFM) and Italian National Institute of Ionizing Radiation Metrology (ENEA-INMRI). *Phys Med* 2025;131:104925.

ARTICLE IN PRESS

44. Lehmann J, Hussein M, Barry MA, et al. SEAFARER - A new concept for validating radiotherapy patient specific QA for clinical trials and clinical practice. *Radiother Oncol.* 2022;171:121-128.
45. Tachibana H, Watanabe Y, Kurokawa S, et al. Multi-institutional study of end-to-end dose delivery quality assurance testing for image-guided brachytherapy using a gel dosimeter. *Brachytherapy* 2022;21:956-967.
46. Mizuno H, Yamashita W, Okuyama H, et al. Dose response of a radiophotoluminescent glass dosimeter for TomoTherapy, CyberKnife, and flattening-filter-free linear accelerator output measurements in dosimetry audit. *Phys Med* 2021;88:91-97.