Physics and Imaging in Radiation Oncology 36 (2025) 100857

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Original Research Article

A multi-institutional dummy run on segmentation variability and plan quality of stereotactic body radiotherapy for oligometastatic disease

Hideaki Hirashima [a,*], Yukinori Matsuo [b], Satoshi Ishikura [c], Mitsuhiro Nakamura [d], Ikuno Nishibuchi [e], Daisuke Kawahara [e], Yoshihisa Shimada [f], Yoshiro Nakahara [g], Teiji Nishio [h], Naoto Shikama [i], Shun-ichi Watanabe [j], Isamu Okamoto [k], Toshiyuki Ishiba [l], Fumikata Hara [m], Tadahiko Shien [n], Takashi Mizowaki [a], Lung Cancer Surgical Study Group (LCSSG)the Lung Cancer Study Group (LCSG), Breast Cancer Study Group (BCSG) and Radiation Therapy Study Group (RTSG) of the Japan Clinical Oncology Group (JCOG)

[a] *Department of Radiation Oncology and Image-Applied Therapy, Graduate School of Medicine, Kyoto University, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan*
[b] *Department of Radiation Oncology, Kindai University Faculty of Medicine, Osaka 589-8511, Japan*
[c] *Department of Radiation Oncology, St. Luke's International Hospital, St. Luke's International University, Tokyo, Japan*
[d] *Department of Advanced Medical Physics, Graduate School of Medicine, Kyoto University, Kyoto, Japan*
[e] *Department of Radiation Oncology, Graduate School of Biomedical Health Sciences, Hiroshima University, Hiroshima, Japan*
[f] *Department of Surgery, Tokyo Medical University, Tokyo, Japan*
[g] *Department of Respiratory Medicine, Kitasato University Hospital, Kanagawa, Japan*
[h] *Medical Physics Laboratory, Division of Health Science, Graduate School of Medicine, The University of Osaka, Osaka, Japan*
[i] *Department of Radiation Oncology, Juntendo University Graduate School of Medicine, Tokyo, Japan*
[j] *Department of Thoracic Surgery, National Cancer Center Hospital, Tokyo, Japan*
[k] *Department of Respiratory Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan*
[l] *Department of Breast Surgery, Institute of Science Tokyo, Tokyo, Japan*
[m] *Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya City, Japan*
[n] *Department of Breast and Endocrine Surgery, Okayama University Hospital, Okayama, Japan*

ARTICLE INFO

Keywords:
Oligometastatic disease
Dummy run
Segmentation variability
Dose distribution accuracy

ABSTRACT

Background and purpose: Oligometastatic disease represents limited metastatic burden, and local ablative therapies such as stereotactic body radiotherapy (SBRT) may improve survival. However, inter-institutional variability in target segmentation and treatment planning can compromise treatment quality. This study aimed to evaluate the segmentation variability and dose distribution quality of SBRT in oligometastatic settings using a multi-institutional dummy run approach.

Methods and materials: Sixty-nine institutions were provided with two anonymized cases of adrenal and spine metastases to delineate targets and organs at risk (OARs) and create intensity-modulated radiotherapy plans following a protocol. Variability was quantified using the Dice similarity coefficient (DSC), Hausdorff distance, and mean distance to agreement. Plan qualities were assessed using the Paddick conformity index, modified gradient index, and a new three-dimensional conformity–gradient index (3D-CGI). Knowledge-based planning (KBP) was applied to explore potential improvements in OAR sparing.

Results: All submitted plans met protocol dose constraints. However, substantial segmentation variability was observed, particularly for the spine case. Among 136 plans, 79% demonstrated acceptable conformity and dose gradients, with 3D-CGI < 6 correlating with favorable distributions. Mean DSC was 0.93 for the clinical target

Abbreviations: SBRT, stereotactic body radiotherapy; OARS, organs at risk; IMRT, intensity-modulated radiotherapy; 3D-CGI, three-dimensional conformity-gradient index; KBP, knowledge-based planning; RTQA, radiotherapy quality assurance; ROI, region of interest; CT, computed tomography; MRI, magnetic resonance image; PTV, planning target volume; GTV, gross tumor volume; PRV, planning risk volume; PCI, Paddick conformity index; MGI, modified gradient index; DSC, Dice similarity coefficient; DVH, dose-volume histogram; SD, standard deviation; CTV, clinical target volume.

* Corresponding author.
E-mail address: hhideaki@kuhp.kyoto-u.ac.jp (H. Hirashima).

https://doi.org/10.1016/j.phro.2025.100857
Received 11 July 2025; Received in revised form 23 October 2025; Accepted 24 October 2025
Available online 1 November 2025

volume and 0.76 for the cauda equina, which showed the highest variability. KBP reduced OAR doses for the adrenal case but showed limited impact for the spine case.
Conclusions: Although dose constraints were achieved, segmentation variability remained substantial, particularly for the cauda equina in the spine case. These findings emphasize inter-institutional differences and the need for standardization and tools to improve SBRT consistency.

1. Introduction

Oligometastases are tumors with only a few metastases to organs beyond the primary site [1,2]. This term describes the limited distant metastases or recurrences that allow for local therapy aimed at a small number of lesions. Local curative therapy combined with systemic treatment may improve long-term survival [3]. Phase III clinical trials have explored local therapies for various cancers, considering the unique factors of each type [4]. Several phase III trials have been initiated worldwide to investigate the role of curative local therapy for oligometastatic disease, including trials evaluating oligo-recurrence after surgical resection of non–small cell lung cancer and local therapy for oligometastatic breast cancer [5,6]. Adequate technology or techniques (e.g., surgery, stereotactic body radiotherapy [SBRT], brachytherapy) are required to treat oligometastatic diseases with curative intent [7–13]. SBRT is a well-established option in selected oligometastatic settings [7]. Given the increasing use of SBRT, ensuring its accuracy and consistency across institutions requires rigorous radiotherapy quality assurance (RTQA) [14,15].

In the RTQA program, assessing contours and treatment planning in multi-institutional trials helps reduce variation [16,17]. RTQA has also decreased the variation in treatment quality in clinical trials [18]. Institutions participating in multi-institutional clinical trials are generally required to meet specific credentialing criteria before patient enrollment. These typically include prior experience in SBRT for relevant treatment sites, verification of treatment delivery accuracy through RTQA, and submission of a dummy run for central review to standardize intensity-modulated radiotherapy (IMRT) planning.

This study aimed to evaluate the segmentation variability and dose distribution quality of SBRT in oligometastatic settings using a multi-institutional dummy run approach. This dummy run offers insights into contouring and planning at participating institutions. The ultimate aim was to standardize IMRT planning for adrenal and spine metastases.

2. Materials and methods

2.1. Dummy run including segmentation and treatment planning

A joint dummy run was performed for the Japan Clinical Oncology Group (JCOG) studies JCOG2108 (Oligo-R) and JCOG2110 (OLIGAMI) [5,6] using two anonymized cases (adrenal and spine metastases) from the first author's institution to evaluate curative local therapy for oligometastatic disease. CT datasets were distributed to 69 institutions. Following the JCOG2108 protocol, 136 treatment plans (69 adrenal, 67 spine) were submitted for central review by the JCOG2108/2110 WG.

The clinical trial JCOG2108 was approved by the Institutional Ethical Review Board of the National Cancer Center Hospital East Certified Review Board (approval number: CRB3180009) and conducted in accordance with the ethical standards stipulated in the 1964 Declaration of Helsinki and its later amendments.

2.2. Segmentation instructions

The target segmentation protocol is summarized in Supplemental **Table S1**. The region of interest (ROI; planning target volume [PTV] + 5 cm) was created by extending all PTVs by 5 cm, and organs at risk (OARs) included in the ROI were targeted for contour extraction. For parallel organs (e.g., lungs, liver, kidneys), the entire organ was contoured to evaluate dose–volume parameters. For serial organs (e.g., spinal cord, cauda equina, esophagus), only the relevant segment within PTV + 5 cm was contoured, as maximum dose evaluation is crucial.

Adrenal case: A breath-hold CT was acquired and provided to all institutions as the planning dataset. The target was the left adrenal tumor. The gross tumor volume (GTV) was obtained from the JCOG2108WG. As breath-hold delivery was reproducible, no additional internal margin was applied. According to the protocol, each institution defined the PTV for adrenal lesions and OARs, including the stomach, duodenum, small bowel, large bowel, kidneys, ureter, spinal canal, and aorta.

Spine case: T2-weighted MRI and CT were acquired and provided to all institutions. The patient had metastasis to the second lumbar vertebra. The GTV was provided by JCOG2108WG. According to the protocol, each institution contoured the clinical target volume (CTV), PTV, and OARs, including the duodenum, small bowel, large bowel, kidneys, ureter, cauda equina, and aorta. In this study, we defined the upper boundary of the cauda equina, which is the lower end of the spinal cord, as the level between the first and second lumbar vertebrae. During the dummy run, the segmentation was acceptable if the spinal cord and cauda equina were contoured smoothly without gaps. The CTV is defined as the GTV and immediately adjacent spine anatomical compartments at risk of microscopic disease extension, based on the contouring guidelines for spine SBRT [19]. Treatments were required to fall within the per protocol category to ensure uniform coverage.

2.3. Treatment planning instructions

Institutions followed the JCOG2108 protocol according to their treatment machines. Prescribed doses: adrenal, mainly 40 Gy/5 fractions (35 or 45 Gy/5 fractions allowed by preference); spine, 35 Gy/5 fractions. The prescription dose and dose constraints for the PTV and OARs are summarized in Supplemental **Tables S2** and **S3**.

2.4. Rates of per protocol evaluation

The submitted dummy runs were categorized as per protocol, acceptable variation, and unacceptable variation. The review focused on contouring consistency, PTV and planning risk volume (PRV) margins, and whether dose constraints met per protocol or acceptable variation or unacceptable variation criteria. In our analysis, variations in the protocol were categorized as either acceptable or unacceptable. Results were categorized according to Global Harmonization Group [20,21]. Acceptable variations were defined as those that did not affect clinical outcomes but varied according to the protocol's specifications. It included verifying the correct naming conventions for targets and OARs and ensuring complete and accurate target delineation. Unacceptable variations were defined as those that could significantly affect clinical outcomes. Specifically, these included inappropriate target or OAR contouring in terms of location, extent, and dose constraints outside the optimal range. Verification of the PTV and PRV margins was also performed to confirm alignment with the protocol specifications. In addition, the dose constraints acquired from the data of each institution were evaluated.

Conformity and gradient indices were evaluated using data acquired from each institution. The Paddick conformity index (PCI) [22] and modified gradient index (MGI) [23] measure dose conformity around and outside the target, respectively. To assess both conformity and dose

Table 1
Categories of the obtained variations from the protocol in 136 treatment plans for the adrenal and spinal cases from the 69 institutions.

	Adrenal (N = 69)			Spine (N = 67)		
	Total	Unacceptable	Acceptable	Total	Unacceptable	Acceptable
Name of target	17% (N = 12)	0% (N = 0)	17% (N = 12)	17% (N = 12)	0% (N = 0)	17% (N = 12)
Contouring of target	3% (N = 2)	0% (N = 0)	3% (N = 2)	25% (N = 17)	4% (N = 3)	20% (N = 14)
PTV margin	20% (N = 14)	1% (N = 1)	19% (N = 13)	6% (N = 4)	4% (N = 3)	1% (N = 1)
Name of OAR	22% (N = 15)	0% (N = 0)	22% (N = 15)	19% (N = 13)	0% (N = 0)	19% (N = 13)
Contouring of OAR	55% (N = 38)	9% (N = 6)	46% (N = 32)	46% (N = 32)	24% (N = 16)	23% (N = 16)
PRV margin	3% (N = 2)	0% (N = 0)	3% (N = 2)	1% (N = 1)	0% (N = 0)	1% (N = 1)
Dose constraint	0% (N = 0)	0% (N = 0)	0% (N = 0)	0% (N = 0)	0% (N = 0)	0% (N = 0)

gradient, we propose a novel metric, the three-dimensional Conformity-Gradient Index (3D-CGI). The 3D-CGI was calculated as the ratio of MGI to PCI, effectively integrating the coverage and dose gradient into a single quantitative measure. This index was designed such that treatment plans with high PCI and low MGI yielded a lower 3D-CGI value. Because 3D-CGI is highly sensitive to variability in the MGI, it can exhibit a wide range of values. A lower 3D-CGI corresponded to a higher-quality treatment plan, whereas a higher 3D-CGI indicated a lower-quality plan.

2.5. Segmentation variability

Segmentation variability was assessed using the Dice similarity coefficient (DSC), Hausdorff distance (HD), mean distance to agreement (MDA), compared with each institution and the reference contour. The reference segmentation was contoured by an experienced physician at the first author's institution and served as the standard. Variability was evaluated only for ROIs required from all institutions. In the adrenal case, the small bowel and left kidney were assessed, and in the spine case, the CTV, cauda equina, and duodenum. Assessments were limited to a 5 cm boundary around the PTV to focus on clinically significant regions. Inter-observer variability was quantified using the pairwise DSC, HD, and MDA for all unique institutional pairs ($N(N-1)/2$). All indices were calculated with MIM Maestro software (MIM Software Inc., OH, USA).

2.6. Knowledge-based planning review

To perform automated patient-specific QA, we used a knowledge-based planning (KBP) approach (RapidPlan™, Eclipse v16.1, Varian Medical Systems, Palo Alto, CA) and compared the mean multi-institutional DVH with that from the KBP model. KBP-generated plans assessed potential for further optimization, particularly OAR sparing. An overview of the KBP DVH evaluation was as follows:

1) The dose distribution of each institution was imported into the reference contours. Reference contours were generated by a physician at the first author's institution.
2) New plans for the adrenal and spine cases were optimized with the KBP model, developed from 69 multi-institutional treatment plans.
3) The mean DVH from the multi-institutional data on the reference contour was compared with that from the KBP model. These plans provided feedback and helped establish consistent radiotherapy standards across centers. For the adrenal case, we evaluated small bowel, duodenum, left kidney, and D2 cm, and for the spine case, we evaluated cauda equina, small bowel, duodenum, and D2 cm.

3. Results

For the adrenal case, 17% institutions showed inconsistent target naming and 3% lacked an internal target volume. Unacceptable and acceptable PTV margin variations were observed in 1% and 19% institutions, respectively. OAR delineation issues occurred in 22% (incorrect naming) and 55% (guideline variation) institutions, whereas PRV margin variation occurred in 3%. No submissions exceeded the dose constraints (Tables 1 and 2). For the spine case, 17% plans showed inconsistent target naming. Unacceptable CTV contouring variation and acceptable sectoral target variation occurred in 4% and 20% institutions, respectively. Unacceptable OAR delineation variation, especially involving cauda equina and bowel, occurred in 24% institutions. PRV margin variation was observed in only 1% institutions. Supplemental **Table S4** summarizes the plan characteristics. For the adrenal case, 67 of the 69 institutions prescribed 40 Gy/5 fractions, one prescribed 45 Gy/5 fractions, and one prescribed 35 Gy/5 fractions. Despite varied device combinations, no plan exceeded dose constraints. There was no correlation between dose indices and planning parameters

Table 2
Gathering pitfalls for the adrenal and spine cases among 136 treatment plans from the 69 institutions through the dummy run test.

Item	Disease site	Variation	Contents
Name of target	Adrenal	Acceptable	The ROI name is incorrect.
	Spine	Acceptable	The ROI name is incorrect.
Contouring of target	Adrenal	Acceptable	A CTV(ITV) has been assigned, but it is not properly defined.
	Spine	Unacceptable	The spinal CTV did not cover the entire vertebral body and was incomplete.
		Acceptable	The spinal CTV was set wider than the sector specified in the protocol.
PTV margin	Adrenal	Unacceptable	The PTV margin does not follow the 2-mm protocol requirement.
		Acceptable	The PTV margin is smaller than the reported value.
	Spine	Unacceptable	The PTV margin exceeds the 2-mm protocol requirement (e.g., set to 3 mm).
		Acceptable	The PTV is contoured as though it were the ROI minus the PRV.
Name of OAR	Adrenal	Acceptable	The ROI name is incorrect.
	Spine	Acceptable	The ROI name is incorrect.
Contouring of OAR	Adrenal	Unacceptable	The small bowel is contoured in the wrong location.
		Acceptable	Not all OAR ROIs within the specified range were delineated.
	Spine	Unacceptable	The spinal cord and cauda equina are contoured in the wrong location.
		Acceptable	Not all OAR ROIs within the specified range were delineated.
PRV margin	Adrenal	Acceptable	The PRV for the spinal canal is not set.
	Spine	Acceptable	The PRV for the spinal cord is not set.
Dose constraint			None

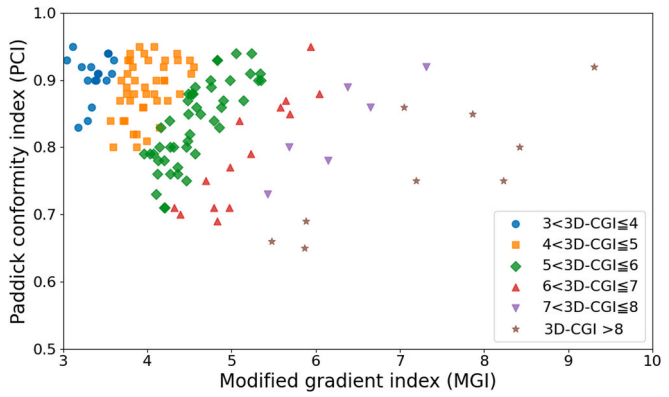

Fig. 1. Variability in conformity metrics, including the Paddick conformity index and modified gradient index with three-dimensional conformity-gradient index for adrenal and spine plans.

(machine, energy, algorithm) or between target and OAR doses. The lower dose quartile in almost all critical organs was found to be below "per protocol" thresholds. Subgroup analysis showed protocol deviations across all institution types (university hospitals, cancer centers, national hospital facilities, community hospitals), indicating they were not dependent on facility scale (Supplemental **Fig. S1**).

Mean ± standard deviation (SD) values for the PCI, MGI, and 3D-CGI were 0.85 ± 0.07, 5.0 ± 1.2, and 5.9 ± 1.6, respectively, for the adrenal case and 0.85 ± 0.07, 4.1 ± 0.7, and 4.8 ± 1.2, respectively, for the spine case. Supplemental **Tables S5** and **S6** and Supplemental **Fig. S2** summarize the results of other dose conformity indices along with those employed in the present study; no differences were observed. When the 3D-CGI was < 6, the PCI remained at ≥ 0.7 and the MGI was ≤ 5.5, regardless of the treatment site (Fig. 1 and Supplemental **Fig. S3**). Among the 136 plans, 79% (n = 107) demonstrated acceptable conformity and dose gradients, with 3D-CGI values < 6 correlating with favorable dose distributions.

Fig. 2 and Supplemental **Fig. S4** illustrates the segmentations from 69 submissions for the adrenal and spine cases overlaid with the reference contour. Means ± SDs for DSC, HD, and MDA for the small bowel and left kidney were 0.82 ± 0.11 and 0.91 ± 0.02, 17.0 ± 9.5 mm and 9.6 ± 3.4 mm, and 1.6 ± 2.1 mm and 0.95 ± 0.02 mm, respectively, for the adrenal case, while those for the CTV, cauda equina, and duodenum were 0.93 ± 0.05, 5.8 ± 5.8 mm, and 0.75 ± 0.69 mm; 0.76 ± 0.17, 17.5 ± 15.9 mm, and 2.5 ± 2.1 mm; and 0.86 ± 0.14, 15.9 ± 18.6 mm, and 2.2 ± 4.4 mm, respectively, for the spine case (**Table 3**). Pairwise DSC, HD, and MDA for all ROIs are in Table 3. Variability patterns between approaches were consistent, although absolute values differed due to the comparison framework. The cauda equina showed the lowest pairwise DSC, indicating difficulty in defining its anatomical extent and substantial inter-institution variability. The CTV showed the highest agreement and smallest geometric deviations across institutions.

Fig. 3 and Supplemental **Fig. S5** compare the multi-institutional mean DVHs with those generated using the KBP model. For the adrenal case, the KBP model successfully reduced the medium-to-low dose range delivered to the OAR. In contrast, for the spine case, the KBP model did not yield a clear reduction in the OAR dose.

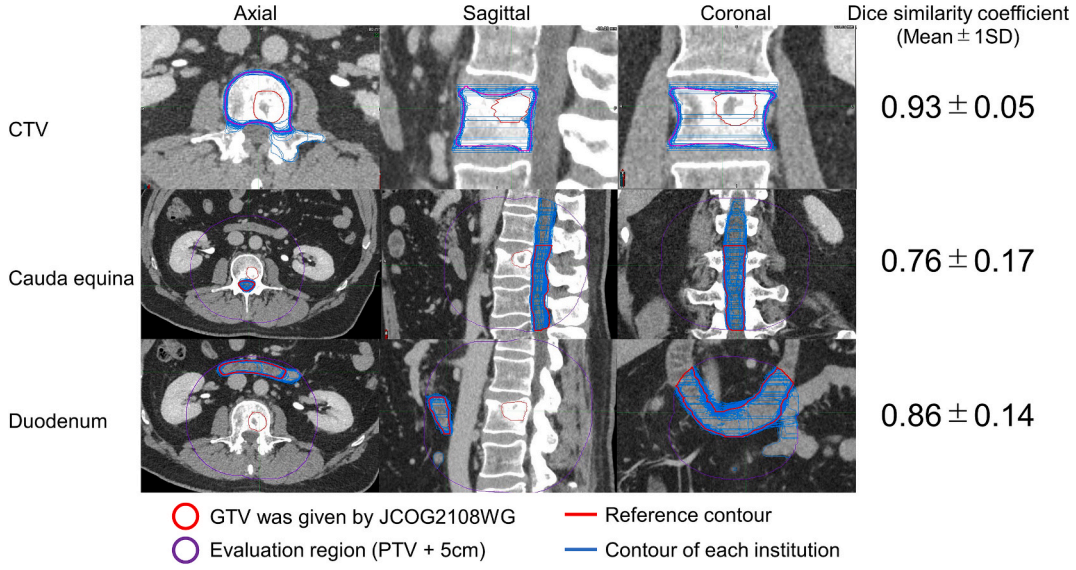

Fig. 2. Multi-institutional segmentation for clinical target volume, cauda equina, and duodenum in the spine case. The reference contour is shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Segmentation accuracy for the adrenal and spine cases. The mean value ± standard deviation for reference-based and pairwise-based DSC, HD and MDA are shown.

Disease sites	Organs	DSC		HD [mm]		MDA [mm]	
		Reference	Pairwise	Reference	Pairwise	Reference	Pairwise
Adrenal	Bowel Small	0.82 ± 0.11	0.75 ± 0.18	17.0 ± 9.5	19.4 ± 13.9	1.6 ± 2.1	2.7 ± 3.3
	Kidney Left	0.91 ± 0.02	0.91 ± 0.03	9.6 ± 3.4	11.5 ± 3.8	1.0 ± 0.0	1.0 ± 0.4
Spine (L2)	CTV	0.93 ± 0.05	0.91 ± 0.06	5.8 ± 5.8	7.1 ± 6.1	0.8 ± 0.7	1.0 ± 0.8
	Cauda equina	0.76 ± 0.17	0.67 ± 0.19	17.5 ± 15.9	22.8 ± 15.9	2.5 ± 2.1	3.3 ± 2.3
	Duodenum	0.86 ± 0.14	0.78 ± 0.18	15.9 ± 18.6	24.0 ± 21.1	2.2 ± 4.4	3.7 ± 6.0

Abbreviations: DSC, Dice similarity coefficient; HD, Hausdorff distance; MDA, mean distance to agreement.

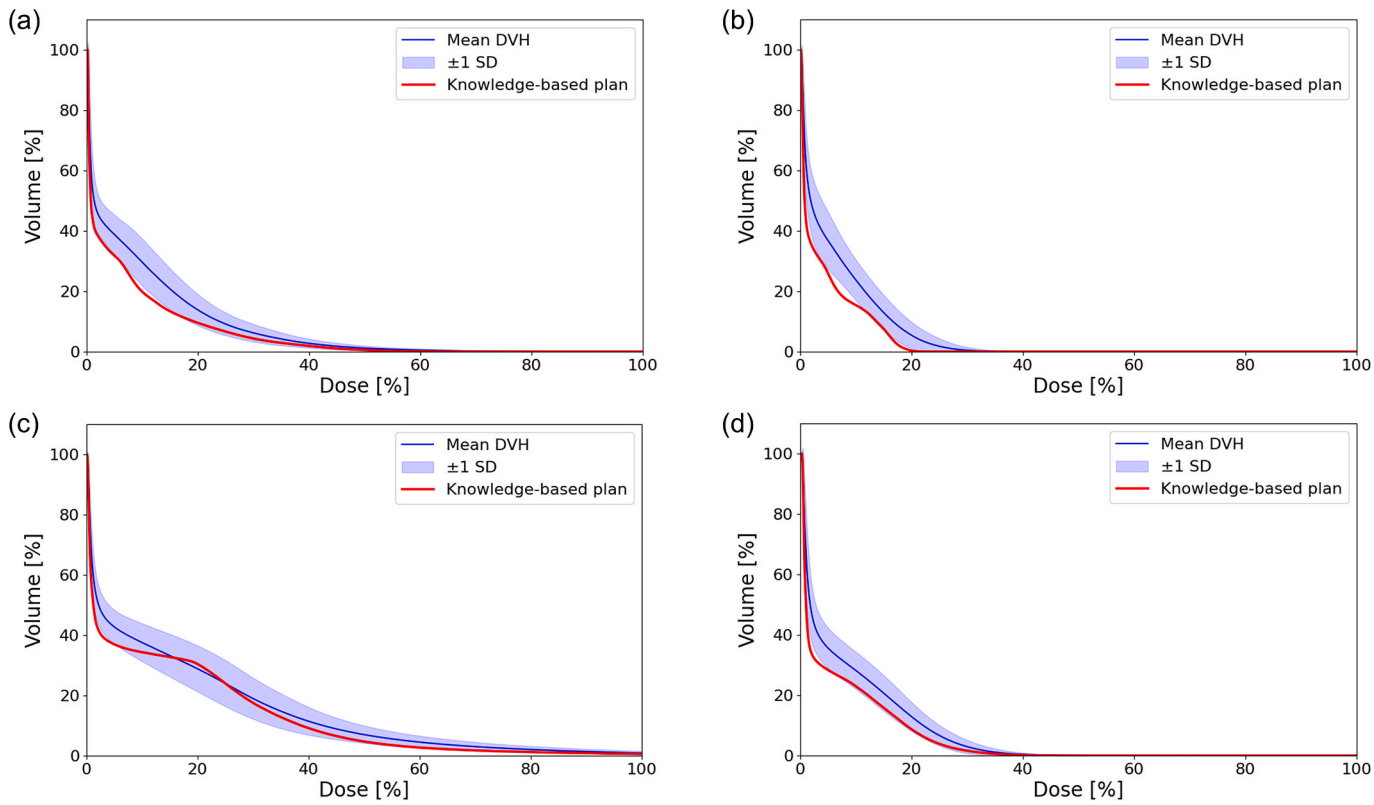

Fig. 3. Mean dose-volume histogram (DVH) (± 1 standard deviation) and knowledge-based plan DVH for (a) small bowel, (b) large bowel, (c) left kidney, and (d) D2 cm in the adrenal case.

4. Discussion

This study presents key technical considerations and quality indicators for SBRT planning in oligometastases and demonstrates the value of KBP-based feedback. Although dose constraints were met, dummy runs from 69 institutions revealed considerable variability in target and OAR contouring, especially for the spine. These findings underscore the need for precise delineation of targets and critical structures to optimize SBRT efficacy and minimize toxicity. Consistent 3D-CGI values < 6 indicate effective planning and highlight KBP's role as a feedback tool to promote standardized, high-quality planning.

Ensuring treatments meet protocol or acceptable variation levels is essential for trial integrity, as quality variation can affect outcomes [14,24–26]. In this study, the most frequent protocol variations involved incorrect OAR delineation and CTV contouring errors in the spine case. These issues stem from limited SBRT experience across organs and insufficient familiarity with consensus guidelines. Lung SBRT is well established in Japan, with the JCOG1408 being the only nationwide clinical trial [27]. SBRT for other organs often depends on institutional experience, leading to variability in quality. Spinal SBRT, in particular, remains less common. Although this study followed international contouring guidelines, some institutions may have relied on locally developed protocols. Meanwhile, JCOG2211, a phase III trial on re-irradiation SBRT for painful spinal metastases, follows similar contouring guidelines despite differences in dose fractions [17,28]. Given these challenges, protocol adherence, systematic feedback, standardized training, and evidence-based guidelines are essential to unify SBRT, promote adoption, improve consistency, and enhance care. This approach can promote broader adoption of SBRT, improve treatment consistency, and enhance patient care.

In SBRT planning, both dose conformity and dose falloff steepness must be considered along with established dose indices [23]. However, these factors have received little attention for the treatment of oligometastases at various anatomical sites. Our results matched previous studies, with **supplementary** analyses confirming favorable outcomes across metrics. We also compared 3D-CGI with other conformity and gradient indices from clinical trials (Supplementary **Tables S5** and **S6** and Supplemental **Fig. S2**). Previous studies have mainly focused on lung SBRT and different fractionation schemes. Our findings confirm that similar dose characteristics apply to other treatment sites. Based on this, we propose the 3D-CGI as a new index for evaluating both dose concentration and steepness. These parameters are usually assessed separately, giving an incomplete view of plan quality (Fig. 1). We found that a 3D-CGI value of ≤ 6 met these requirements (PCI > 0.7 and MGI $<$ 5.5). This integrated framework offers a more comprehensive and quantitative assessment of the plan quality.

In SBRT clinical trials, accurate contouring is essential because it directly affects the clinical outcomes. Dose distributions are generated from contours and optimized to meet constraints, but inaccurate contouring may still satisfy protocol criteria. Therefore, relying solely on institutional dose metrics for plan evaluation may be insufficient. Given the steep dose gradients observed in this study, even small undercontoured CTV segments beyond the PTV margin can rapidly fall below the prescribed dose. This underscores the need for accurate contouring and supports integrating contour review with targeted dosimetric checks in QA. A quantitative assessment of the contouring variability, along with structured feedback, is essential. This study found the greatest inter-institutional variation in cauda equina delineation for the spine case. Previous research shows cauda equina contouring is challenging, underscoring the need for standardized training and protocols to ensure reproducibility and reliability [29]. To address this issue, feedback and discussions on correct delineation methods have been conducted to improve consistency. This approach enhances our understanding of trial complexity and promotes standardized treatment strategies. Several previous reports have assessed contouring accuracy using dummy run analysis in clinical trials [30–35]. They emphasize the

importance of precise contouring and education regarding high-precision treatments. Additionally, several studies have investigated automated contouring models based on clinical trial data and demonstrated their effectiveness [36,37]. In the future, leveraging dummy run data to develop and distribute such models can help reduce the segmentation variability across institutions.

Visual assessment alone often cannot confirm OAR sparing due to patient-specific anatomy. Comparing institutional plans with the KBP provides objective benchmarks, highlighting opportunities for further dose optimization. This confirms that KBP is a practical feedback tool for improving clinical trial outcomes [37–43]. Systematic guidance, such as the KBP, helps institutions reduce variability in OAR contouring and improve overall plan quality. Open challenges such as Open-KBP and Auto-RTP mainly benchmark automated planning approaches [44,45]. In contrast, our study evaluated inter-institutional variability in contouring and planning within a clinical trial QA framework. We used KBP primarily to provide achievable planning goals, not as a head-to-head comparison of automated methods.

This study has several limitations. First, it focused only on adrenal and spine SBRT, limiting assessment of other sites. Institutions with less experience may struggle to achieve consistent planning quality for untreated sites without specific feedback. Therefore, individual clinical reviews should be conducted for actual cases to provide ongoing evaluation and feedback. Second, this study is limited by the generalizability of the KBP model. The KBP models constructed in this dummy run were based specifically on adrenal and L2 spine metastases, demonstrating their utility at these anatomical sites. However, because of the limited variety of organs represented in the training dataset, the model may not be applicable to other anatomical sites commonly treated for oligometastatic diseases. Therefore, the effectiveness of the KBP-based planning may be restricted when applied to anatomical regions that are not adequately represented in the training data. Third, only two cases, both obtained from the same institution, were used for the dummy run. This limited number and origin of cases may not fully represent the diversity of patient anatomy or institutional practices, which could further restrict the generalizability of our findings. However, similar QA dummy run studies also used one or two cases, making this a common, accepted limitation [30–35].

In conclusion, this multi-institutional dummy run study evaluated SBRT planning for oligometastases. Although dose constraints were met, notable segmentation variability, especially in cauda equina for the spine case, was observed. These findings provide evidence of inter-institutional variability in contouring and dosimetry and demonstrate KBP's value for improvement. The results highlight the need for stronger standardization, feedback, and education. Future efforts should focus on refining guidelines and bolstering training to ensure consistent and high-quality SBRT and more predictable patient outcomes.

Funding

None.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank all the staff involved in radiation therapy at our institution. We would like to express our sincere gratitude to all the institutional staff who helped us. We also thank the members of the JCOG Medical Physics Working Group of the Radiation Therapy Study Group for their support and suggestions. This research was partly supported by JSPS KAKENHI Grant Number JP24K18826.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2025.100857.

References

[1] Hellman S, Weichselbaum RR. Oligometastases. J Clin Oncol 1995;13:8–10. https://doi.org/10.1200/JCO.1995.13.1.8.
[2] Punglia RS, Morrow M, Winer EP, Harris JR. Local therapy and survival in breast cancer. N Engl J Med 2007;356:2399–405. https://doi.org/10.1056/NEJMra065241.
[3] Yamamoto T, Niibe Y, Aoki M, Shintani T, Yamada K, Kobayashi M, et al. Analyses of the local control of pulmonary oligometastases after stereotactic body radiotherapy and the impact of local control on survival. BMC Cancer 2020;20:997. https://doi.org/10.1186/s12885-020-07514-9.
[4] Sogono P, Ball DL, Siva S. SABR-COMET: a new paradigm of care lights up the twilight of metastatic disease. Ann Transl Med 2019;7:615. https://doi.org/10.21037/atm.2019.10.96.
[5] Shimada Y. Oligo-recurrence in lung cancer; the most curable state among advanced disease? Cancers (Basel) 2024;16:4086. https://doi.org/10.3390/cancers16234086.
[6] Ishiba T, Nishibuchi I, Hara F, Sasaki K, Sadate R, Sekino Y, et al. Abstract PO3-19-01: a randomized controlled trial of metastasis-directed therapy for oligometastases in breast cancer (OLIGAMI trial; JCOG2110). Cancer Res 2024;84:PO3–19-01-PO3-19-01. https://doi.org/10.1158/1538-7445.SABCS23-PO3-19-01.
[7] Lievens Y, Guckenberger M, Gomez D, Hoyer M, Iyengar P, Kindts I, et al. Defining oligometastatic disease from a radiation oncology perspective: an ESTRO-ASTRO consensus document. Radiother Oncol 2020;148:157–66. https://doi.org/10.1016/j.radonc.2020.04.003.
[8] Bilski M, Korab K, Stąpór-Fudzińska M, Ponikowska J, Brzozowska A, Sroka Ł, et al. HDR brachytherapy versus robotic-based and linac-based stereotactic ablative body radiotherapy in the treatment of liver metastases—A dosimetric comparison study of three radioablative techniques. Clin Transl Radiat Oncol 2024;48:100815. https://doi.org/10.1016/j.ctro.2024.100815.
[9] Bilski M, Peszyńska-Piorun M, Konat-Bąska K, Brzozowska A, Korab K, Wojtyna E, et al. Radiotherapy as a metastasis directed therapy for liver oligometastases—comparative analysis between CT-guided interstitial HDR brachytherapy and two SBRT modalities performed on double-layer and single layer LINACs. Front Oncol 2024;14:1478872. https://doi.org/10.3389/fonc.2024.1478872.
[10] Cisek P, Bilski M, Ponikowska J, Wojtyna E, Fijuth J, Kuncman Ł. EORTC/ESTRO defined induced oligopersistence of liver metastases from colorectal cancer—outcomes and toxicity profile of computer tomography guided high-dose-rate brachytherapy. Clin Exp Metastasis 2025;42:29. https://doi.org/10.1007/s10585-025-10348-z.
[11] Dejonckheere CS, Bilski M, Nour Y, Scafa D, Cisek P, Korab K, et al. A dosimetric comparison of different radiotherapy modalities for non-resected oligometastatic liver Disease. Clin Transl Radiat Oncol 2025;52:100947. https://doi.org/10.1016/j.ctro.2025.100947.
[12] Bilski M, Korab K, Orzechowska M, Ponikowska J, Cisek P, Jereczek-Fossa BA, et al. Comprehensive cohort study: computer tomography-guided high-dose rate brachytherapy as metastasis-directed therapy for liver metastases from colorectal cancer in repeat oligoprogression. Radiol Med 2025;130:694–705. https://doi.org/10.1007/s11547-025-01988-y.
[13] Cisek P, Kuncman Ł, Jereczek-Fossa BA, Wojtyna E, Orzechowska M, Sroka S, et al. Computed tomography guided high dose rate brachytherapy for induced oligoprogression of colorectal cancer liver metastases. Sci Rep 2025;15:22735. https://doi.org/10.1038/s41598-025-09227-0.
[14] Brooks C, Miles E, Hoskin PJ. Radiotherapy trial quality assurance processes: a systematic review. Lancet Oncol 2024;25:e104–13. https://doi.org/10.1016/S1470-2045(23)00625-3.
[15] Corrigan KL, Kry S, Howell RM, Kouzy R, Jaoude JA, Patel RR, et al. The radiotherapy quality assurance gap among phase III cancer clinical trials. Radiother Oncol 2022;166:51–7. https://doi.org/10.1016/j.radonc.2021.11.018.
[16] Okamoto H, Murakami N, Isohashi F, Kasamatsu T, Hasumi Y, Iijima K, et al. Dummy-run for standardizing plan quality of intensity-modulated radiotherapy for postoperative uterine cervical cancer: Japan Clinical Oncology Group study (JCOG1402). Radiat Oncol 2019;14:133. https://doi.org/10.1186/s13014-019-1340-y.
[17] Ito K, Nakajima Y, Nakamura N, Ito Y, Nishio T, Mizowaki T, et al. Contouring compliance and variability of targets and organs at risk in spine stereotactic body Radiotherapy: a 34-Institution study. Radiother Oncol 2025;207:110877. https://doi.org/10.1016/j.radonc.2025.110877.
[18] Okamoto H, Murakami N, Isohashi F, Kasamatsu T, Hasumi Y, Kobayashi H, et al. Plan quality association between dummy run and individual case review in a prospective multi-institutional clinical trial of postoperative cervical cancer patients treated with intensity-modulated radiotherapy: Japan clinical Oncology Group study (JCOG1402). Radiother Oncol 2023;183:109630. https://doi.org/10.1016/j.radonc.2023.109630.
[19] Cox BW, Spratt DE, Lovelock M, Bilsky MH, Lis E, Ryu S, et al. International Spine Radiosurgery Consortium consensus guidelines for target volume definition in spinal stereotactic radiosurgery. Int J Radiat Oncol Biol Phys 2012;83:e597–605. https://doi.org/10.1016/j.ijrobp.2012.03.009.

[20] Melidis C, Bosch WR, Izewska J, Fidarova E, Zubizarreta E, Ulin K, et al. Global harmonization of quality assurance naming conventions in radiation therapy clinical trials. Int J Radiat Oncol Biol Phys 2014;90:1242–9. https://doi.org/10.1016/j.ijrobp.2014.08.348. Erratum. In.
[21] Melidis C, Bosch WR, Izewska J, Fidarova E, Zubizarreta E, Ishikura S, et al. Radiation therapy quality assurance in clinical trials—Global Harmonisation Group. Radiother Oncol 2014;111:327–9. https://doi.org/10.1016/j.radonc.2014.03.023.
[22] Paddick I. A simple scoring ratio to index the conformity of radiosurgical treatment plans. Technical note J Neurosurg 2000;93(Suppl 3):219–22. https://doi.org/10.3171/jns.2000.93.supplement.
[23] Lee J, Dean C, Patel R, Webster G, Eaton DJ. Multi-center evaluation of dose conformity in stereotactic body radiotherapy. Phys Imaging Radiat Oncol 2019;11:41–6. https://doi.org/10.1016/j.phro.2019.08.002.
[24] Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. J Clin Oncol 2010;28:2996–3001. https://doi.org/10.1200/JCO.2009.27.4498.
[25] Brade AM, Wenz F, Koppe F, Lievens Y, San Antonio B, Iscoe NA, et al. Radiation therapy quality assurance (RTQA) of concurrent chemoradiation therapy for locally advanced non-small cell lung cancer in the PROCLAIM phase 3 trial. Int J Radiat Oncol Biol Phys 2018;101:927–34. https://doi.org/10.1016/j.ijrobp.2018.04.015.
[26] Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: a meta-analysis of cooperative group clinical trials. J Natl Cancer Inst 2013;105:387–93. https://doi.org/10.1093/jnci/djt001.
[27] Kimura T, Nagata Y, Eba J, Ozawa S, Ishikura S, Shibata T, et al. A randomized phase III trial of comparing two dose-fractionations stereotactic body radiotherapy (SBRT) for medically inoperable stage IA non-small cell lung cancer or small lung lesions clinically diagnosed as primary lung cancer: Japan Clinical Oncology Group Study JCOG1408 (J-SBRT trial). Jpn J Clin Oncol 2017;47:277–81. https://doi.org/10.1093/jjco/hyw198.
[28] Kita R, Ito K, Machida R, Sekino Y, Nakamura N, Nakajima Y, et al. Randomized phase III study comparing re-irradiation stereotactic body radiotherapy and conventional radiotherapy for painful spinal metastases: Japan Clinical Oncology Group study JCOG2211 (rescore study). Jpn J Clin Oncol 2025;55:189–93. https://doi.org/10.1093/jjco/hyae145.
[29] Dunne EM, Lo SS, Liu MC, Bergman A, Kosztyla R, Chang EL, et al. Thecal sac contouring as a surrogate for the cauda equina and intracanal spinal nerve roots for spine stereotactic body radiation therapy (SBRT): contour variability and recommendations for safe practice. Int J Radiat Oncol Biol Phys 2022;112:114–20. https://doi.org/10.1016/j.ijrobp.2021.08.023.
[30] Gwynne S, Spezi E, Sebag-Montefiore D, Mukherjee S, Miles E, Conibear J, et al. Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case. Br J Radiol 2013;86:20120398. https://doi.org/10.1259/bjr.20120398.
[31] Caravatta L, Macchia G, Mattiucci GC, Sainato A, Cernusco NLV, Mantello G, et al. Inter-observer variability of clinical target volume delineation in radiotherapy treatment of pancreatic cancer: a multi-institutional contouring experience. Radiat Oncol 2014;9:198. https://doi.org/10.1186/1748-717X-9-198.
[32] Fairchild A, Langendijk JA, Nuyts S, Scrase C, Tomsej M, Schuring D, et al. Quality assurance for the EORTC 22071–26071 study: dummy run prospective analysis. Radiat Oncol 2014;9:248. https://doi.org/10.1186/s13014-014-0248-9.
[33] Chung Y, Kim JW, Shin KH, Kim SS, Ahn SJ, Park W, et al. Dummy run of quality assurance program in a phase 3 randomized trial investigating the role of internal mammary lymph node irradiation in breast cancer patients: Korean Radiation Oncology Group 08–06 study. Int J Radiat Oncol Biol Phys 2015;91:419–26. https://doi.org/10.1016/j.ijrobp.2014.10.022.
[34] Lambrecht M, Clementel E, Sonke JJ, Nestle U, Adebahr S, Guckenberger M, et al. Radiotherapy quality assurance of SBRT for patients with centrally located lung tumours within the multicentre phase II EORTC Lungtech trial: benchmark case results. Radiother Oncol 2019;132:63–9. https://doi.org/10.1016/j.radonc.2018.10.025.
[35] Maier SH, Schönecker S, Anagnostatou V, Garny S, Nitschmann A, Fleischmann DF, et al. Dummy run for planning of isotoxic dose-escalated radiation therapy for glioblastoma used in the PRIDE trial (NOA-28; ARO-2024-01; AG-NRO-06). Clin Transl Radiat Oncol 2024;47:100790. https://doi.org/10.1016/j.ctro.2024.100790.
[36] Thor M, Apte A, Haq R, Iyer A, LoCastro E, Deasy JO. Using auto-segmentation to reduce contouring and dose inconsistency in clinical trials: the simulated impact on RTOG 0617. Int J Radiat Oncol Biol Phys 2021;109:1619–26. https://doi.org/10.1016/j.ijrobp.2020.11.011.
[37] Porter EM, Vu C, Sala IM, Guerrero T, Siddiqui ZA. Deep learning for contour quality assurance for RTOG 0933: in-silico evaluation. Radiother Oncol 2024;201:110519. https://doi.org/10.1016/j.radonc.2024.110519.
[38] Xianghua Y, Guo D, Zhao L, Xie C, Zheng D, Yang H, et al. Development and validation of AI delineation of the thoracic RTOG organs at risk with deep learning on multi-institutional datasets. Int Oncol 2025;1:161–71.
[39] Tol JP, Dahele M, Gregoire V, Overgaard J, Slotman BJ, Verbakel WFAR. Analysis of EORTC-1219-DAHANCA-29 trial plans demonstrates the potential of knowledge-based planning to provide patient-specific treatment plan quality assurance. Radiother Oncol 2019;130:75–81. https://doi.org/10.1016/j.radonc.2018.10.005.
[40] Younge KC, Marsh RB, Owen D, Geng H, Xiao Y, Spratt DE, et al. Improving quality and consistency in NRG Oncology Radiation Therapy Oncology Group 0631 for spine radiosurgery via knowledge-based planning. Int J Radiat Oncol Biol Phys 2018;100:1067–74. https://doi.org/10.1016/j.ijrobp.2017.12.276.
[41] Giaddui T, Geng H, Chen Q, Linnemann N, Radden M, Lee NY, et al. Offline quality assurance for intensity modulated radiation therapy treatment plans for NRG-HN001 head and neck clinical trial using knowledge-based planning. Adv Radiat Oncol 2020;5:1342–9. https://doi.org/10.1016/j.adro.2020.05.005.
[42] Hardcastle N, Cook O, Ray X, Moore A, Moore KL, Pryor D, et al. Personalising treatment plan quality review with knowledge-based planning in the TROG 15.03 trial for stereotactic ablative body radiotherapy in primary kidney cancer. Radiat Oncol 2021;16:142. https://doi.org/10.1186/s13014-021-01820-7.
[43] van Gysen K, Kneebone A, Le A, Wu K, Haworth A, Bromley R, et al. Evaluating the utility of knowledge-based planning for clinical trials using the TROG 08.03 post prostatectomy radiation therapy planning data. Phys Imaging. Radiat Oncol 2022;22:91–7. https://doi.org/10.1016/j.phro.2022.05.004.
[44] Babier A, Mahmood R, Zhang B, Alves VGL, Barragán-Montero AM, Beaudry J, et al. OpenKBP-Opt: an international and reproducible evaluation of 76 knowledge-based planning pipelines. Phys Med Biol 2022;67. https://doi.org/10.1088/1361-6560/ac8044.
[45] Gooding MJ, Aluwini S, Guerrero Urbano T, McQuinlan Y, Om D, Staal FHE, et al. Fully automated radiotherapy treatment planning: a scan to plan challenge. Radiother Oncol 2024;200:110513. https://doi.org/10.1016/j.radonc.2024.110513.