Pejman Jabehdar Maralani, MD [*]
Chia-Lin Tseng, MD [‡]
Hamidreza Baharjoo, MD [*]
Erin Wong, MD [*]
Anish Kapadia, MD [*]
Archya Dasgupta, MD [‡]
Peter Howard, MD [*]
Aimee K.M. Chan, MSc [*]
Eshetu G. Atenafu, MSc [§]
Hua Lu, MSc [*]
Pascal Tyrrell, PhD [*¶‖]
Sunit Das, MD, PhD [#]
Seyedmehdi Payabvash, MD [**]
Jay Detsky, MD, PhD [‡]
Zain Husain, MD [‡]
Sten Myrehaug, MD [‡]
Hany Soliman, MD [‡]
Hanbo Chen, MD [‡]
Chris Heyn, MD, PhD [*]
Sean Symons, MD, MBA [*]
Arjun Sahgal, MD [‡]

[*]Department of Medical Imaging, University of Toronto, Toronto, Canada; [‡]Department of Radiation Oncology, University of Toronto, Toronto, Canada; [§]Department of Biostatistics, University Health Network, Toronto, Canada; [¶]Department of Statistical Sciences, University of Toronto, Toronto, Canada; [‖]Institute of Medical Science, University of Toronto, Toronto, Canada; [#]Department of Surgery, Division of Neurosurgery, University of Toronto, Toronto, Canada; [**]Department of Radiology, Yale School of Medicine, New Haven, Connecticut, USA

Correspondence:
Pejman Jabehdar Maralani, MD,
Department of Medical Imaging,
University of Toronto,
2075 Bayview Ave, Room AG270c,
Toronto, Ontario, Canada M4N3M5.
Email: pejman.maralani@utoronto.ca

Received, January 11, 2021.
Accepted, June 9, 2021.
Published Online, August 14, 2021.

The Initial Step Towards Establishing a Quantitative, Magnetic Resonance Imaging-Based Framework for Response Assessment of Spinal Metastases After Stereotactic Body Radiation Therapy

BACKGROUND: There are no established threshold values regarding the degree of growth on imaging when assessing response of spinal metastases treated with stereotactic body radiation therapy (SBRT).

OBJECTIVE: To determine a magnetic resonance imaging-based minimum detectable difference (MDD) in gross tumor volume (GTV) and its association with 1-yr radiation site-specific (RSS) progression-free survival (PFS).

METHODS: GTVs at baseline and first 2 post-SBRT scans (Post1 and Post2, respectively) for 142 spinal segments were contoured, and percentage volume change between scans calculated. One-year RSS PFS was acquired from medical records. The MDD was determined. The MDD was compared against optimal thresholds of GTV changes associated with 1-yr RSS PFS using Youden's J index, and receiver operating characteristic curves between timepoints compared to determine which timeframe had the best association.

RESULTS: A total of 17 of the 142 segments demonstrated progression. The MDD was 10.9%. Baseline-Post2 demonstrated the best performance (area under the curve [AUC] 0.90). Only Baseline-Post2 had an optimal threshold > MDD at 14.7%. Due to large distribution of GTVs, volumes were split into tertiles. Small tumors (GTV < 2 cc) had optimal thresholds of 42.0%, 71.3%, and 37.2% at Baseline-Post1 (AUC 0.81), Baseline-Post2 (AUC 0.89), and Post1-Post2 (AUC 0.77), respectively. Medium tumors (2 ≤ GTV ≤ 8.3 cc) all demonstrated optimal thresholds < MDD, with AUCs ranging from 0.65 to 0.84. Large tumors (GTV > 8.3 cc) had 2 timepoints where optimal thresholds > MDD: Baseline-Post2 (13.3%; AUC 0.97) and Post1-Post2 (11.8%; AUC 0.66). Baseline-Post2 had the best association with RSS PFS for all tertiles.

CONCLUSION: Given a MDD of 10.9%, for small GTVs, larger (>37%) changes were required before local failure could be determined, compared to 11% to 13% for medium/large tumors.

KEY WORDS: Assessment, Disease progression, Metastasis, Outcome, Spine, Stereotactic body radiotherapy

Neurosurgery 89:884–891, 2021 https://doi.org/10.1093/neuros/nyab310 www.neurosurgery-online.com

Stereotactic body radiation therapy (SBRT) is now an established technique for the treatment of painful spinal metastases.

ABBREVIATIONS: GTV, gross tumor volume; **ICC,** intraclass correlation coefficient; **IQR,** interquartile range; **MDD,** minimum detectable difference; **RECIST,** response evaluation criteria in solid tumors; **RSS,** radiation site-specific; **SBRT,** stereotactic body radiation therapy; **SPINO,** Spine response assessment in Neuro-Oncology

Supplemental digital content is available for this article at www.neurosurgery-online.com.

With increasing evidence supporting long-term survival benefits in oligometastatic disease following SBRT,[1,2] including recent evidence from the Canadian Cancer Trials Group-Symptom Control randomized controlled trial (CCTG-SC.24; NCT02512965),[3] the application of spine SBRT will only continue to increase. With this, the need for globally acceptable, quantitative imaging-based response assessment guidelines also continues to grow.

While the response evaluation criteria in solid tumors (RECIST) 1.1 guidelines[4] provides quantifiable definitions for tumor response assessment, its applicability to bone metastases

is limited. In particular, RECIST 1.1 considers only the extraosseous tumor of lytic metastases to be potentially measurable, and sclerotic metastases are considered entirely nonmeasurable. The Spine response assessment In Neuro-Oncology (SPINO) group recognized this limitation and recommended joint assessment by a neuro-radiologist and radiation oncologist with respect to an unequivocal increase in the gross tumor volume (GTV).[5] It was recognized that complicating assessment is the complex anatomy of vertebral segments with relation to the standard orthogonal sagittal and axial planes commonly used in spine magnetic resonance imaging (MRI). With the 3 to 4 mm slice thickness used for routine follow-up spine MRIs,[6] it may not be possible to assess the tumor at exactly the same position on subsequent imaging. Accordingly, evaluation of GTV over multiple slices is the most accurate way to determine tumor size and assess growth or shrinkage.

Under standard of care, spinal metastases that demonstrate growth may require clinical intervention, but even with SPINO's guidance, what constitutes an "unequivocal increase" in GTV remains unknown. Currently no guidelines provide defined, imaging-based thresholds for response assessment in routine clinical care or clinical trials. The first step towards the development of such guidelines is to quantify what is considered a GTV "increase." The goal of this study was to determine what the lowest possible value this could be while accounting for measurement error in spinal metastases treated with SBRT. We secondarily compared this value to how a statistically derived threshold would perform in prognostication of 1-yr radiation site-specific (RSS) progression-free survival (PFS). These data will lay the groundwork to develop quantitative, evidence-based guidelines for response assessment – including a threshold that defines "progression" – in this patient population.

METHODS

Research Ethics Board's approval with waiver of informed consent was obtained for this single-center retrospective study. Data are reported in accordance to STROBE (Strengthening the Reporting of Observational studies in Epidemiology) guidelines for cohort studies.[7]

Patient Selection

All adult (≥18 yr old) patients with de novo spinal metastases treated with 24 Gy in 2 SBRT fractions from January 1, 2009 to December 31, 2015, were included. Other inclusion criteria included availability of the pre-SBRT baseline MRI and 2 routine follow-up spine MRIs, performed at approximately 3 and 6 mo post-SBRT (Post1 and Post2, respectively). Spinal segments were excluded if it was treated with prior radiotherapy, or if the spinal segment had a vertebral compression fracture, decompressive surgery, vertebroplasty or instrumentation, as these interfere with accurate radiologic measurement of GTV.

SBRT and GTV Segmentation

The SBRT workflow and methodology were previously described.[8-10] All MRIs were performed on 2 identical 1.5T GE TwinSpeed Excite scanners (GE Medical Systems, Milwaukee, Wisconsin). All images were anonymized, and GTVs contoured by a single neuro-radiologist with 6 yr of experience using ITK-SNAP (version 3.8.0)[11] on sagittal T1 weighted images (3 mm slice thickness, no gap) (Figure 1). If only axial volumetric T1 imaging were available at baseline, the images were reformatted to the sagittal plane using MIPAV (National Institutes of Health, Bethesda, Maryland) to the same parameters of conventional sagittal T1 imaging. All segmentations were verified by a radiation oncologist with expertise in spine SBRT.

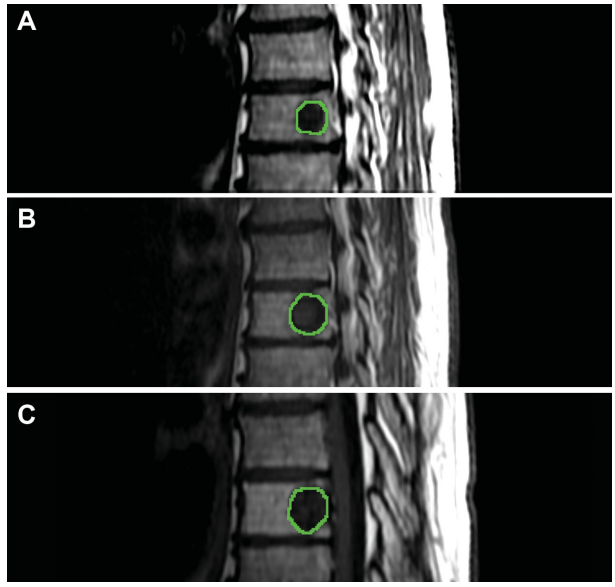

FIGURE 1. *Contoured spinal segment (T8) on baseline* **A**, *first follow-up* **B**, *and second follow-up* **C** *demonstrating progression at the radiated spinal level.*

Intra- and Inter-Rater Reliability of Segmentations

To assess consistency of segmentations, the baseline MRI of 34 spinal segments of varying sizes were randomly chosen, reordered, and recontoured by the neuroradiologist who originally contoured the metastases to determine intrarater reliability. To assess the repeatability of segmentations, 2 additional neuroradiologists also independently contoured the segments to determine inter-rater reliability. All neuroradiologists were blinded to clinical data as well as each other's segmentations during this process.

Clinical Endpoint

The primary outcome was 1-yr RSS PFS, determined by reviewing the electronic medical records from the visit closest to 1-yr follow up. The SPINO criteria[5] were used to determine RSS PFS, with progression defined as: unequivocal increase in GTV, new or progressive epidural tumor, or neurological deterioration attributable to pre-existing epidural disease. Outcomes were classified as either progression or nonprogression.

Statistical Analyses

Each spinal segment was evaluated independently. GTV changes were calculated as percent changes between scans, using the planning MRI as

the baseline scan. To determine the intrarater reliability, the intraclass correlation coefficient (ICC) was calculated using k = 3 while inter-rater reliability was calculated using k = 2. To determine a threshold which could be used to define a change in GTV beyond any measurement error, the minimum detectable difference (MDD, Equation 1)[12] was calculated based on the baseline and Post1 GTVs, which were segmented by 1 neuroradiologist, twice:

$$\mathrm{MDD} = z\,\mathrm{score}_{\text{level of confidence}} \times \mathrm{SD}_{|\text{percent change}|} \times \sqrt{(2\,[1 - r_{\text{test -retest}}])}, \quad (1)$$

where the z-score was 1.96, associated with 95% confidence interval; $\mathrm{SD}_{|\text{percent change}|}$ was the standard deviation of the absolute percent change in GTV; and $r_{\text{test-retest}}$ was the coefficient of test-retest reliability, estimated from the ICC.

Although the MDD can be applied in both directions, we have applied it in the positive direction only, given the clinical applications for determining a value for GTV "increase," as opposed to "decrease." In practice, when applying the MDD, GTV changes above the MDD can be considered as real changes while GTV changes below the MDD may be due to measurement error.

The area under the receiver operator characteristic curve (AUC) was then calculated to assess the discriminative ability of the MDD for predicting 1-yr RSS PFS. To further assess its utility, we compared the MDD's performance against the optimal thresholds derived from Youden's J index, which gives the highest sum of sensitivity and specificity. Univariate logistic regression was performed to associate percent changes in GTV at the various time points (Baseline-Post1, Baseline-Post2, and Post1-Post2) with 1-yr RSS PFS.

All statistical analyses were performed using R version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria). Anonymized data will be held in accordance with institutional policies. To request access to data, please contact the corresponding author.

RESULTS

The initial database contained 284 treated spinal segments from 150 consecutive patients. A total of 22 segments were excluded due to vertebroplasty, decompressive surgery, or vertebral compression fracture; 14 segments were excluded as the tumor extended beyond the MRI field of view; and 106 segments had <2 follow-up scans. Therefore, a total of 142 spinal segments from 85 patients were included (Table 1).

A summary of the baseline characteristics of the spinal segments from baseline to Post2 are presented in Table 1. Median follow-up was 30.5 mo (interquartile range [IQR], 13.9-56.1 mo), though analyses were limited to 1-yr RSS PFS. At 1 yr, 19 spinal segments demonstrated progression. Out of these, 2 were initially believed to represent progression, but subsequent scans confirmed pseudoprogression.[8,13-15] As this is a transient phenomenon that mimics disease progression on imaging but otherwise resolves without therapeutic intervention, these were subsequently considered not progressed, therefore the final cohort had 17 progression cases. A version of the results, where the 2 pseudoprogression cases were not included in the analyses, can be found in **Supplemental Tables 1 and 2** and **Supplemental Figure**.

High ICC was observed for intrarater reliability (0.99, $P < .001$, 95% CI 0.98-0.995) and inter-rater reliability (0.995, $P < .001$, 95% CI 0.992-0.997). These strengthen the confidence of the results for the MDD, which was determined to be 10.9%.

TABLE 1. Baseline Characteristics and Treatment Response of All 142 Spinal Segments Included in the Study

Variable	
Primary cancer site	N (%)
Breast	57 (40.1)
Colorectal	8 (5.6)
Nonsmall cell lung carcinoma	20 (14.1)
Melanoma	2 (1.4)
Prostate	23 (16.2)
Renal cell carcinoma	23 (16.2)
Squamous cell carcinoma	3 (2.1)
Other*	4 (2.8)
Unknown	2 (1.4)
Spine metastatic location	N (%)
Cervical	9 (6.3)
Thoracic	83 (58.5)
Lumbar	36 (25.3)
Sacral	14 (9.9)
Epidural disease at the treated segment[17]	N (%)
0	105 (73.9)
1a	15 (10.6)
1b	16 (11.3)
1c	5 (3.5)
2	1 (0.7)
3	0 (0)
Spinal metastases tumor type	N (%)
Sclerotic	52 (36.7)
Lytic	66 (46.5)
Mixed	24 (16.9)
Treatment	N (%)
24 Gy/2 fractions	142 (100.0)
Time between scans	Median (IQR) in days
Post1	89 (74-103) days after baseline
Post2	91 (67-107) days after Post1
Contoured GTV sizes	Median (IQR) in cc
Baseline	5.6 (1.2-10.2)
Post1	4.4 (1.0-10.6)
Post2	4.0 (0.9-8.9)
1-yr RSS PFS	N (%)
No progression	125 (88.0)
Progression	17 (12.0)
Small tumors (GTV < 2 cc)	5/17 (29.4)
Medium tumors (2 ≤ GTV ≤ 8.3 cc)	6/17 (35.3)
Large tumors (GTV > 8.3 cc)	6/17 (35.3)

Abbreviations: Gy, Gray; IQR, interquartile range; GTV, gross tumor volume; RSS PFS, radiation site-specific progression-free survival.
*Primaries were uterine, hepatocellular carcinoma, and neuroendocrine tumors.

Global Findings

The optimal threshold values derived from Youden's J index, and their associated AUC, sensitivity, and specificity in relation to 1-yr RSS PFS are presented in Table 2. Only the threshold between the Baseline-Post2 scan was above the MDD. For other timepoints, when using MDD as the threshold (Table 3), the sensitivity decreased at all timepoints while the specificity

TABLE 2. The Threshold and Its Associated AUC, Sensitivity, and Specificity in Determining 1-Yr RSS PFS

	Percentage change between baseline and first post-treatment follow-up				Percentage change between baseline and second post-treatment follow-up				Percentage change between first and second follow-ups			
	Threshold (%)	AUC	Sensitivity (%)	Specificity (%)	Threshold (%)	AUC	Sensitivity (%)	Specificity (%)	Threshold (%)	AUC	Sensitivity (%)	Specificity (%)
All tumors (n = 142)	−0.22	0.81	76.5	75.2	14.7	0.90	76.5	93.6	9.43	0.7	52.9	90.4
Small tumors (GTV < 2cc; n = 47)	42	0.81	60	100	71.3	0.89	80	97.6	37.2	0.77	60	90.5
Medium tumors (2 ≤ GTV ≤ 8.3 cc; n = 47)	−0.26	0.76	66.7	78.0	0.79	0.84	83.3	75.6	9.4	0.65	50	92.7
Large tumors (GTV > 8.3 cc; n = 48)	0.64	0.89	100	73.8	13.3	0.97	100	95.2	11.8	0.66	50	100

Abbreviations: AUC, associated area under the receiver operator curve; GTV, gross tumor volume.

TABLE 3. The Sensitivity and Specificity When Thresholds, Originally Below the Value of the MDD, Are Set to the MDD (10.9%)

	Percentage change between baseline and first post-treatment follow-up			Percentage change between baseline and second post-treatment follow-up			Percentage change between first and second follow-ups		
	Threshold (%)	Sensitivity (%)	Specificity (%)	Threshold (%)	Sensitivity (%)	Specificity (%)	Threshold (%)	Sensitivity (%)	Specificity (%)
All tumors (n = 142)	10.9	47.1	93.6	14.7 (*Unchanged*)	76.5	93.6	10.9	47.1	90.4
Small tumors (GTV < 2cc; n = 47)	42 (*unchanged*)	60	100	71.3 (*unchanged*)	80	97.6	37.2 (*unchanged*)	60	90.5
Medium tumors (2 ≤ GTV ≤ 8.3 cc; n = 47)	10.9	33.3	92.7	10.9	50	92.7	10.9	33.3	92.7
Large tumors (GTV > 8.3 cc; n = 48)	10.9	50	92.9	13.3 (*unchanged*)	100	95.2	11.8 (*unchanged*)	50	100

Abbreviations: GTV, gross tumor volume.

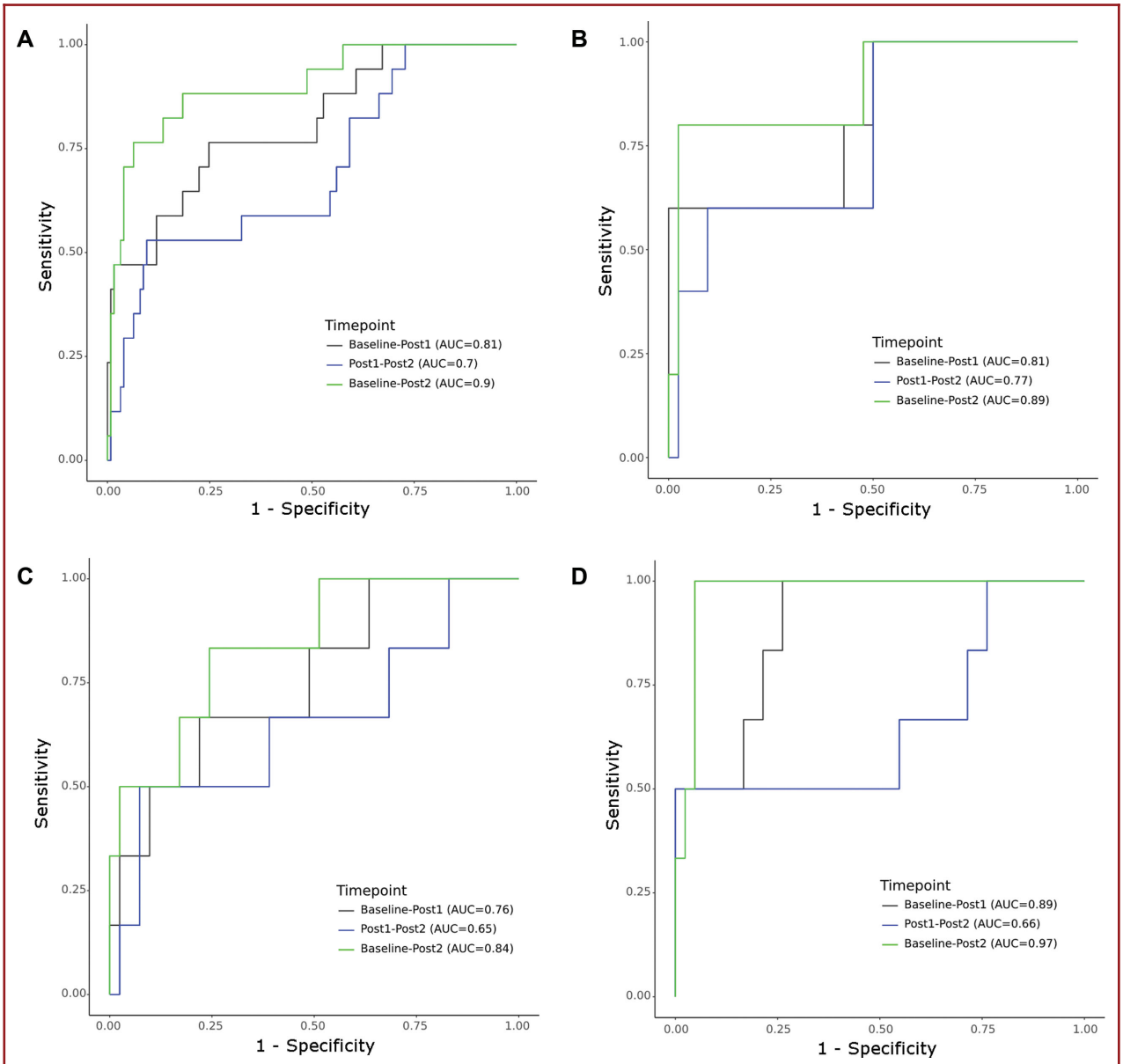

FIGURE 2. *The AUCs of the different timepoints studied in the prognostication of 1-yr RSS PFS for all tumors* **A**, *and when stratified into small* **B** *(GTV < 2 cc), medium* **C** *(2 ≤ GTV ≤ 8.3 cc), and large* **D** *(GTV > 8.3 cc) tumors.*

remained the same or increased (all >90%) compared to the optimal values derived from Youden's J index.

Figure 2A demonstrates the AUCs used for predicting 1-yr RSS PFS. The highest performing AUC was from Baseline-Post2 (AUC 0.90), followed by Baseline-Post1 (AUC 0.81), and Post1-Post2 (AUC 0.7). Baseline-Post2 had a significantly higher discriminative ability than Post1-Post2 ($P = .004$) while univariate analysis comparing other permutations of time points did not reveal any statistical significance (Table 4).

Tertiles

Due to the large distribution of GTVs (range, 0.034-44.5 cc; IQR, 1.2-10.2 cc), data were split based on the baseline GTV into tertiles. Tumors with baseline GTVs less than 2 cc are hereafter referred to as "small"; those with GTVs greater than 8.3 cc as "large"; and those with GTVs in between these values as "medium." The 17 progressed segments included 5 (29.4%) small, 6 (35.3%) medium, and 6 (35.3%) large tumors. A summary of the thresholds, their associated AUC, sensitivity,

TABLE 4. Discriminative Ability of Each Timepoint's AUC Relative to the 1-Yr RSS PFS Status

	Baseline-Post1 vs Baseline-Post2	Baseline-Post1 vs Post1-Post2	Baseline-Post2 vs Post1-Post2
All tumors (n = 142)	0.10	0.30	0.004
Small tumors (GTV < 2cc; n = 47)	0.40	0.78	0.21
Medium tumors (2 ≤ GTV ≤ 8.3 cc; n = 47)	0.53	0.66	0.18
Large tumors (GTV > 8.3 cc; n = 48)	0.17	0.27	0.054

and specificity with regards to 1-yr RSS PFS are presented in Tables 2 and 3.

For small tumors, the optimal threshold for predicting 1-yr RSS PFS was above the MDD at all times points and ranged from 37.2% to 71.3%, with AUCs ranging from 0.77 to 0.89, without any statistical difference when comparing the various time points ($P > .05$; Figure 2B and Table 4).

Medium-sized tumors had optimal thresholds below the MDD at all timepoints, ranging from −0.26% to 9.4%, with AUCs ranging from 0.65 to 0.84. When the threshold was set to the MDD, specificity at all timepoints increased to 92.7% but sensitivity declined to 33.3% to 50%. No statistical differences were detected between the AUCs of using the various time points ($P > .05$; Figure 2C and Table 4).

Large tumors demonstrated optimal thresholds close to the MDD, except for Baseline-Post1 where the threshold was 0.64%. Using the MDD at this timeframe resulted in lower sensitivity (100% to 50%) but higher specificity (73.8% to 92.9%). The change between Baseline-Post2 had the highest discriminative ability of all tertiles and timepoints, with an AUC of 0.97. No statistical differences were detected between the AUCs of any timepoint ($P > .05$; Figure 2D and Table 4).

The best performing AUC was observed when comparing the Baseline-Post2 scans for all tumor sizes, followed by Baseline-Post1, then Post1-Post2, although no time point was significantly better than the other.

DISCUSSION

Key Results

Despite widespread emergence of SBRT, imaging assessment of tumor response for spinal metastases is challenging even with the RECIST 1.1[4] and SPINO[5,16] guidelines. Here, we present the first step towards the development of quantitative, imaging-based guidelines by identifying the MDD, and how it compares against a statistically derived, optimal threshold. Furthermore, the associations of specific timepoints with 1-yr RSS PFS highlight the potential utility that such guidelines could provide.

Generalizability

The development of an imaging-based assessment guideline, specific to spinal metastases treated with SBRT, is crucial with the establishment of SBRT. While RECIST 1.1[4] includes guidelines on the measurability of bone lesions, unique limitations exist when applying these towards spinal metastases. For example, under RECIST 1.1, purely lytic and mixed lytic-sclerotic metastases are considered "measurable" only if there is an extraosseous soft tissue component ≥ 10 mm; sclerotic metastases are considered nonmeasurable. Most spinal metastases have Bilsky grade 0 to 1 epidural disease,[17] as such the GTVs that are entirely intraosseous (Bilsky 0) would be considered nonmeasurable under RECIST 1.1. Even if the tumor has a Bilsky 1 epidural component, this is almost invariably a small fraction of the GTV. Moreover, the anteroposterior dimension (ie, thickness) of the epidural component is shorter than both transverse or craniocaudal dimensions, and very often less than 10mm, rendering it again nonmeasurable. Tumor measurements are further complicated by the complex 3-dimensional shape of the vertebral bodies and their posterior elements, making it very difficult to capture the largest dimension of the tumor. Lastly, due to time constraints, it is not feasible to acquire volumetric follow-up imaging with the same parameters as treatment planning MRI.[5]

To begin addressing some of these issues, our study quantified what minimum change in volume within the first 2 follow-up MRIs, when therapeutic interventions may still have a meaningful impact on outcome, should be considered as truly increased.

Interpretation

Without size stratification, the thresholds for optimizing sensitivity and specificity were only above the MDD of 10.9% when assessing the changes between Baseline-Post2. However, the negative threshold for Baseline-Post1 highlights the need to account for initial size of the tumor when determining a clinically meaningful volumetric change. For example, a threshold of −0.22% (indicating shrinkage) between Baseline-Post1 had the best combination of sensitivity and specificity for 1-yr RSS PFS. However, as this threshold is below the MDD, changes within this range may be due to measurement error. Therefore, while further work is required to identify an appropriate threshold for this timeframe, the MDD of 10.9% can be used as the smallest clinically meaningful increase in GTV even if it does not yield the highest AUC, until a value for "progression" is defined.

For small tumors (<2 cc), large changes in GTV ranging from 37% to 71% had the highest combination of sensitivity and specificity for progression. This is unsurprising, as the absolute size

change for "small" tumors may be clinically irrelevant despite large percentage changes; this is also reflected in the high specificity, but comparatively low sensitivity of the thresholds. In medium-sized tumors ($2 \leq GTV \leq 8.3$ cc), the threshold between all timepoints was below the MDD. Although the thresholds were derived statistically, again this indicates that the reliability of the thresholds for tumors of this size may be low and therefore 10.9% should be used as the minimum threshold to determine "growth." For particularly large tumors (>8.3 cc), a GTV increase of 13.3% had the highest sensitivity and specificity. Since it is also greater than the MDD, it could, therefore, be used for risk prediction.

These data highlight the clinical significance of the second post-SBRT scan. While it is possible that Post2 had better discrimination than Post1 due to being closer in time to the outcome studied, the high sensitivity and specificity for the various derived thresholds support the use of this time point. Clinically, it is also early enough that further intervention could be considered.

Limitations and Future Directions

As with all retrospective studies, selection bias could not be fully accounted for. Due to the complex nature of spinal metastases, there is inherent subjectivity with contouring; however, this was limited by having one neuroradiologist contour all GTVs with subsequent verification by a radiation oncologist, and high inter-rater reliability. Lastly, application of the MDD which was derived from the entire dataset to stratified data assumes that the MDD is not affected by tumor size. Despite the inclusion of 142 spinal segments, this was necessary as there was insufficient power to derive tertile-specific MDDs. In future work, inclusion of the CCTG-SC.24 multi-institutional dataset[3] (NCT02512965) will also help increase power to determine and investigate tertile-specific MDDs, as well as the effect of other parameters such as histology, tumor type, and systemic disease and chemotherapy.

CONCLUSION

The results from this data-driven study represent the first step towards the development of objective imaging-based response assessment criteria for spinal metastases treated with SBRT. We identified an MDD of 10.9%, which accounts for measurement error and represents the minimum to define a GTV "increase." Furthermore, we identified the second post-SBRT scan as a possible point for clinical action; when the MDD is assessed in this timeframe, there was reasonable performance in prognostication. More work is needed before a threshold value to define "progression" is derived.

Funding

This study did not receive any funding or financial support.

Disclosures

The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article. Dr Sahgal reports other from Abbvie, grants and other from Elekta/Elekta AB, other from Accuray Inc, other from Varian Medical Systems, other from BrainLAB, other from Merck, other from Roche, other from International Stereotactic Radiosurgery Society, other from Medtronic Kyphon, and other from VieCue, outside the submitted work. Dr Chia-Lin Tseng has received travel accommodations/expenses and honoraria for past educational seminars from Elekta and belongs to the Elekta MR-linac Research Consortium. Dr Husain reports other from Elekta, outside the submitted work. Dr Myrehaug reports personal fees from Advanced Accelerator Applications, personal fees from Novartis, personal fees from Ipsen, outside the submitted work.

REFERENCES

1. Chen H, Louie AV, Higginson DS, Palma DA, Colaco R, Sahgal A. Stereotactic radiosurgery and stereotactic body radiotherapy in the management of oligometastatic disease. *Clin Oncol (R Coll Radiol)*. 2020;32(11):713-727.
2. Palma DA, Olson RA, Harrow S, et al. Stereotactic ablative radiotherapy for the comprehensive treatment of oligometastatic cancers: long-term results of the SABR-COMET randomized trial. *Int J Radiat Oncol Biol Phys*. 2020;108(3):S88-S89.
3. Sahgal A, Myrehaug SD, Siva S, et al. Stereotactic body radiotherapy versus conventional external beam radiotherapy in patients with painful spinal metastases: an open-label, multicentre, randomised, controlled, phase 2/3 trial. *Lancet Oncol*. 2021;22(7):P1023-1033.
4. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228-247.
5. Thibault I, Chang EL, Sheehan J, et al. Response assessment after stereotactic body radiotherapy for spinal metastasis: a report from the SPIne response assessment in Neuro-Oncology (SPINO) group. *Lancet Oncol*. 2015;16(16):e595-e603.
6. American College of Radiology. ACR–ASNR–SCBT-MR–SSR practice parameter for the performance of magnetic resonance imaging (MRI) of the adult spine. Practice Parameters. Revised 2018 (Resolution 19).
7. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12(12):1495-1499.
8. Jabehdar Maralani P, Winger K, Symons S, et al. Incidence and time of onset of osseous pseudoprogression in patients with metastatic spine disease from renal cell or prostate carcinoma after treatment with stereotactic body radiation therapy. *Neurosurgery*. 2019;84(3):647-654.
9. Tseng CL, Soliman H, Myrehaug S, et al. Imaging-based outcomes for 24 Gy in 2 daily fractions for patients with de novo spinal metastases treated with spine stereotactic body radiation therapy (SBRT). *Int J Radiat Oncol Biol Phys*. 2018;102(3):499-507.
10. Cox BW, Spratt DE, Lovelock M, et al. International Spine Radiosurgery Consortium consensus guidelines for target volume definition in spinal stereotactic radiosurgery. *Int J Radiat Oncol Biol Phys*. 2012;83(5):e597-e605.
11. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116-1128.
12. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther*. 2006;86(5):735-743.
13. Taylor DR, Weaver JA. Tumor pseudoprogression of spinal metastasis after radiosurgery: a novel concept and case reports. *J Neurosurg Spine*. 2015;22(5):534-539.
14. Bahig H, Simard D, Letourneau L, et al. A study of pseudoprogression after spine stereotactic body radiation therapy. *Int J Radiat Oncol Biol Phys*. 2016;96(4):848-856.
15. Amini B, Beaman CB, Madewell JE, et al. Osseous pseudoprogression in vertebral bodies treated with stereotactic radiosurgery: a secondary analysis of prospective phase I/II clinical trials. *AJNR Am J Neuroradiol*. 2016;37(2):387-392.
16. Laufer I, Lo SS, Chang EL, et al. Population description and clinical response assessment for spinal metastases: part 2 of the SPIne response assessment in Neuro-Oncology (SPINO) group report. *Neuro-oncology*. 2018;20(9):1215-1224.
17. Bilsky MH, Laufer I, Fourney DR, et al. Reliability analysis of the epidural spinal cord compression scale. *J Neurosurg Spine*. 2010;13(3):324-328.

Supplemental digital content *is available for this article at www.neurosurgery-online.com.*

Supplemental Table 1. The equivalent of the article's Table 2, when cases of pseudoprogression (n = 2) were excluded from analysis.

Supplemental Table 2. The equivalent of the article's Table 3, when cases of pseudoprogression (n = 2) were excluded from analysis.

Supplemental Figure. The AUCs of the different timepoints studied in the prognostication of 1-yr RSS PFS for all tumors **A**, and when stratified into small **B** (GTV < 2.2 cc), medium **C** (2.2 ≤ GTV ≤ 8.4 cc), and large **D** (GTV > 8.4 cc) tumors, with exclusion of 2 pseudoprogression cases. Relative to the AUCs when the pseudoprogression cases were included, AUCs are the same or improved, with the exception of small tumors at Baseline-Post1, which has slightly less discriminative ability (0.80 vs 0.81 when pseudoprogression cases are included).