Nodular Leptomeningeal Disease—A Distinct Pattern of Recurrence After Postresection Stereotactic Radiosurgery for Brain Metastases: A Multi-institutional Study of Interobserver Reliability

Brandon E. Turner, MD, MSc[*], Roshan S. Prabhu, MD[†,‡], Stuart H. Burri, MD[†,‡], Paul D. Brown, MD[§], Erqi L. Pollom, MD, MS[*], Michael T. Milano, MD, PhD[‖], Stephanie E. Weiss, MD[#], Michael Iv, MD[**], Nancy Fischbein, MD[**], Hany Soliman, MD[††], Simon S. Lo, MD[§§], Samuel T. Chao, MD[‖‖], Brett W. Cox, MD[***], James D. Murphy, MD, MS[†††], Gordon Li, MD[‡‡], Melanie Hayden Gephart, MD, MAS[‡‡], Seema Nagpal, MD[§§§], Banu Atalar, MD[‖‖‖], Melissa Azoulay, MD, MSc[###], Reena Thomas, MD, PhD[§§§], Gayle Tillman, MD[****], Ben Y. Durkee, MD, PhD[††††], Jennifer L. Shah, MD[‡‡‡‡], Scott G. Soltys, MD[*]

[*]Department of Radiation Oncology, Stanford Cancer Institute, Stanford, California

[†]Southeast Radiation Oncology Group, Charlotte, North Carolina

[‡]Levine Cancer Institute, Atrium Health, Charlotte, North Carolina

[§]Department of Radiation Oncology, Mayo Clinic, Rochester, Minnesota

[‖]University of Rochester Medical Center, Rochester, New York

[#]Fox Chase Cancer Center, Philadelphia, Pennsylvania

[**]Department of Neuroimaging and Neurointervention, Stanford University, Stanford, California

[††]Department of Radiation Oncology, Sunnybrook Odette Cancer Centre, University of Toronto, Toronto, ON, Canada

[§§]Department of Radiation Oncology, University of Washington, Seattle, Washington

[‖‖]Department of Radiation Oncology, Cleveland Clinic, Cleveland, Ohio

[***]Department of Radiation Medicine, Northwell Health, New York, New York

[†††]Department of Radiation Medicine and Applied Sciences, University of California San Diego, San Diego, California

[‡‡]Department of Neurosurgery, Stanford School of Medicine, Stanford, California

[§§§]Department of Neurology, Stanford University, Stanford, California

||||||Department of Radiation Oncology, Acibadem University School of Medicine, Istanbul, Turkey

###Department of Radiation Oncology, McGill University Health Center, Montreal, Canada

****Department of Radiation Oncology, Massachusetts General Hospital, Boston, Massachusetts

††††Department of Radiation Oncology, SwedishAmerican, Rockford, Illinois

‡‡‡‡Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan

Abstract

Purpose: For brain metastases, surgical resection with postoperative stereotactic radiosurgery is an emerging standard of care. Postoperative cavity stereotactic radiosurgery is associated with a specific, underrecognized pattern of intracranial recurrence, herein termed *nodular leptomeningeal disease* (nLMD), which is distinct from classical leptomeningeal disease. We hypothesized that there is poor consensus regarding the definition of LMD, and that a formal, self-guided training module will improve interrater reliability (IRR) and validity in diagnosing LMD.

Methods and Materials: Twenty-two physicians at 16 institutions, including 15 physicians with central nervous system expertise, completed a 2-phase survey that included magnetic resonance imaging and treatment information for 30 patients. In the "pretraining" phase, physicians labeled cases using 3 patterns of recurrence commonly reported in prospective studies: local recurrence (LR), distant parenchymal recurrence (DR), and LMD. After a self-directed training module, participating physicians completed the "posttraining" phase and relabeled the 30 cases using the 4 following labels: LR, DR, classical leptomeningeal disease, and nLMD.

Results: IRR increased 34% after training (Fleiss' Kappa $K = 0.41$ to $K = 0.55$, $P < .001$). IRR increased most among noncentral nervous system specialists (+58%, $P < .001$). Before training, IRR was lowest for LMD ($K = 0.33$). After training, IRR increased across all recurrence subgroups and increased most for LMD (+67%). After training, 27% of cases initially labeled LR or DR were later recognized as nLMD.

Conclusions: This study highlights the large degree of inconsistency among clinicians in recognizing nLMD. Our findings demonstrate that a brief self-guided training module distinguishing nLMD can significantly improve IRR across all patterns of recurrence, and particularly in nLMD. To optimize outcomes reporting, prospective trials in brain metastases should incorporate central imaging review and investigator training.

Introduction

Brain metastases are the most common malignant brain tumors, with a median survival of 1.6 to 17.1 months after diagnosis.[1] Historically, after surgical resection of a brain metastases, adjuvant whole brain radiation therapy (WBRT) was the standard irradiation technique, affording improved intracranial tumor control, although not significantly affecting survival.[2,3] To minimize the neurocognitive side effects of WBRT, many institutions have reported outcomes of stereotactic radiosurgery (SRS) to only the surgical cavity in retrospective[4–6] and single-arm prospective series.[7,8] The utilization of postoperative SRS will undoubtedly increase given the recent positive results of the cooperative group N107C phase III trial.[9] This trial randomized patients with up to 4 brain metastases, of which 1

was resected, to postoperative WBRT versus SRS. Cognitive deterioration at 6 months was significantly reduced in patients receiving postoperative SRS (52%) versus WBRT (85%), with no difference in overall survival.

The omission of WBRT after surgical resection of brain metastases is associated with a shift in the pattern of intracranial recurrence. Multiple studies[9–12] report a higher rate of leptomeningeal disease (LMD), synonymously termed *leptomeningeal carcinomatosis*, after postoperative SRS compared with WBRT. This increased rate is presumably related to iatrogenic dissemination of tumor cells at the time of resection. Because these cells outside the cavity are not treated with postoperative SRS, they later recur as LMD when WBRT is omitted.

The true incidence of postoperative LMD is unclear, ranging from 8% to 35% across studies.[10,13] The reason for this wide range is multifactorial, with potential differences due to tumor histology,[14,15] location,[15,16] size,[17] and pial involvement,[7] as well as type of surgical resection.[16] Additionally, differences in imaging follow-up and underreporting of LMD are potential factors. For example, the single-institution postresection SRS prospective trial from M D Anderson Cancer Center noted a 28% rate of LMD compared with 7% on the multi-institutional N107C trial.[9,18,19] This discrepancy in reported LMD outcomes weakens the development of accurate treatment recommendations and suggests underlying discordance in physicians' assessment (and consequent treatment) of LMD.

LMD diagnosis is often made using a combination of neuroimaging and clinical suspicion, without obtaining CSF. The radiographic signs typically associated with LMD, for which we propose the term *classical leptomeningeal disease (cLMD)*, include (1) enhancement of the cranial nerves, cisterns, cerebellar folia, and sulci[20]; and (2) *Zuckerguss* or diffuse "sugar-coating" enhancement across the surface of the brain.[21–25] Cagney et al[26] recently reported an 11% incidence of pachymeningeal seeding of the dural or outer arachnoid after surgical resection of a brain metastasis compared with 0% without resection. In this report, the authors chose the anatomic term *pachymeningeal seeding* to contrast it from *leptomeningeal disease* of the subarachnoid to distinguish the types of CSF spread of tumor. With a similar thought to recent reports[27] on postradiosurgery toxicity preferring the term *adverse radiation effect* as the imaging correlate of histologically defined radiation necrosis, we have chosen the terms *classical LMD* (cLMD) versus *nodular LMD* (nLMD),[28] because these outcomes are defined by imaging rather than histology. We hypothesize that intracranial nLMD, which occurs predominantly in the postresection setting with postoperative SRS[28] as opposed to the nLMD seen in the spine ("drop metastases") as noted in the EANO guidelines,[29] is morphologically distinct, underrecognized, and commonly misreported as distal intraparenchymal recurrence (DR).

We propose a formal description of nLMD (Table 1). We have created a brief, self-guided training module to distinguish nLMD from cLMD and DR. To assess the baseline level of agreement in LMD diagnosis, we tasked a set of physicians (including many with expertise in central nervous system [CNS] oncology) with reviewing follow-up neuroimaging from patients who received postoperative SRS and labeling each case based on their assessment of the patient's pattern of intracranial recurrence. To assess the benefit of our nLMD training

module, each physician subsequently followed the self-guided training module before repeating a similar task in labeling patterns of recurrence represented in patients' follow-up neuroimaging. We hypothesize that a training module that standardized the definition of LMD would improve interobserver variability in reporting of intracranial patterns of recurrence. Additionally, we explore how outcomes differ between physicians self-identified to be specialists in CNS tumors compared with non-CNS physicians.

Methods and Materials

Patients and imaging

With Stanford University institutional review board approval, we retrospectively identified 30 patients with brain metastases treated with SRS to a postresection cavity who developed intracranial recurrence on follow-up imaging. Patients had both gadolinium-enhanced T1-weighted 2D spin echo and 3D inversion recovery-spoiled gradient recalled echo images for assessment. In addition to extracting representative magnetic resonance imaging (MRI), we obtained a screen capture of patients' SRS plan targeting the resection cavity. We selected cases that displayed a single pattern of intracranial recurrence, excluding those with multiple patterns of intracranial recurrence (eg, both local cavity recurrence and distal parenchymal recurrence were excluded). We also excluded patients with adverse radiation effect (ie, the radiographic correlate of radiation necrosis) or with MRI studies of poor technical quality. Images were anonymized and exported with MIM 6.7.11 (MIM Software Inc, Cleveland, OH). We provided survey respondents with an electronic package of MRI DICOM files for each patient and MIMviewer 3.4.14 (MIM Software Inc.), a lightweight viewing software for medical imaging.

Training module creation

We identified patients with imaging features typical of LMD per prior classification[24,25] and the authors' previous experience treating patients with multiple forms of LMD. These features were organized into a 12-page training document along with representative images to illustrate key features that we propose distinguish nLMD from cLMD and DR (see document overview in Table 1 and a sample image in Fig. 1). The complete training document is available online (File A, available at https://doi.org/10.1016/j.ijrobp.2019.10.002).

Survey creation

We collected survey data using REDCap electronic data capture tools hosted at our institutional center.[30] We created 2 survey arms within REDCap: "pretraining" and "posttraining" arms. Each arm featured the same cases but with order (and anonymized patient labels) randomized. In each survey arm for each case we provided respondents with: (1) a single slice from the patient's preresection MRI that demonstrated the position of the original brain metastasis; (2) a 3-plane extract from the patient's SRS treatment plan that displayed the surgical cavity and radiation isodose curves; and (3) up to 3 (axial, coronal, sagittal) MRI images, viewable in MIMviewer, demonstrating the study pattern of intracranial recurrence (see representative example in Fig. E1, available at https://doi.org/10.1016/j.ijrobp.2019.10.002).

In the "pretraining" arm, we asked respondents to review the information and imaging for each case and then label which type of intracranial recurrence was present using the 3 patterns of recurrence commonly reported in prospective trials: (1) local cavity recurrence (LR); (2) DR; and (3) leptomeningeal disease (Fig. 2). After assigning a label, raters were also asked to rate their confidence that the label was accurate on a 5-point Likert scale from 0 (least confidence) to 4 (greatest confidence).

In the "posttraining" arm, respondents repeated this procedure but with the order and anonymized patient IDs randomized again. In the posttraining arm, raters used one of the following 4 labels: (1) local recurrence; (2) distal parenchymal recurrence; (3) cLMD; and (4) nLMD. Rater confidence was assessed again in the same manner. In both the "pretraining" and "posttraining" sessions, respondents were instructed that no cases represented adverse radiation effect.

Survey respondents

In March 2018, 22 physicians participated in the study and all completed both parts of the survey. Respondents consisted of radiation oncologists (n = 16; 73%), neuro-oncologists (n = 2; 9%), neurosurgeons (n = 2; 9%), and neuroradiologists (n = 2; 9%) from 16 different institutions. Physicians included those described as CNS specialists (n = 15; 68%) and radiation oncologists without a CNS-specific clinical practice (n = 7; 32%). Physicians with CNS expertise were identified based on (1) a record of publications on CNS oncology in academic medical journals and (2) a clinical appointment with predominantly CNS or patients with cancer.

Statistical analysis

We assessed percentage agreement for each arm by measuring the proportion of cases for which all raters assigned a case the same label. We calculated Fleiss' Kappa statistic (K), which has a 0 to 1 scale, to assess the degree of agreement above the agreement one could attribute to random chance across multiple raters and multiple cases.[31] Subgroup analysis was also performed using Fleiss' Kappa to determine the degree of agreement within specific patterns of intracranial recurrence. Although there is no universally accepted interpretation of Kappa statistics, the study by Landis and Koch[32] is the most commonly cited. However, we avoided an absolute interpretation of the Kappa statistic using such benchmarks given their widespread criticism for being inherently arbitrary, difficult to translate across domains, and unresponsive to differences in the implications of disagreement.[33,34]

We tested the effect of assigned labels on confidence levels using 1-way between-subjects ANOVA. Posthoc comparisons were performed using the Tukey HSD test for multiple comparisons.[35]

We assessed for differences between subgroups of raters using Wilcoxon's test (see supplemental methods, available at https://doi.org/10.1016/j.ijrobp.2019.10.002, for details of subgroup comparisons). We used Cramer's V (V) to calculate the association between categorical data and Pearson's correlation (r) for numeric data.

All analyses were performed using the R statistical programming language, version 3.5.0 (accessed at r-project.org).

Results

Interrater agreement and reassignment of case labels

Interrater reliability improved with the training module, with the most improvement seen among non-CNS specialists. We compared rater agreement before and after taking the self-guided training module. Overall, raters were more likely to agree on a case's assigned label after the training, with overall interrater reliability (IRR) improving by +34% ($K = 0.41$ to $K = 0.55$, $P < .001$), and with an even greater improvement among non-CNS specialists (IRR + 58%, $P < .001$).

At the group-wide consensus level, for each case, the most frequently assigned label across all CNS raters (the "collective label") seldom differed before and after training. The pretraining LMD label was considered equivalent to either nLMD or cLMD posttraining. For CNS-specialized physicians, the consensus label was equivalent for 29 out of 30 cases. In contrast, non-CNS specialists' labels were more influenced by the training. Non-CNS specialists' consensus labels changed more frequently and were equivalent for only 21 out of 30 cases.

A set of labels was also assigned by the senior author. Although these do not necessarily constitute a gold standard, the senior author's labels were made with the benefit of the patients' physical examination, supporting diagnostic tests, and knowledge of the patients' ultimate clinical course. CNS raters' consensus labels were identical to our senior author's labeling in 29 out of 30 cases both pre- and posttraining (Cramer's $V = 0.94$ and $V = 0.95$ respectively). Non-CNS raters' pre- and posttraining consensus labels were identical to the senior author's labeling in only 22 ($V = 0.69$) and 24 ($V = 0.71$) cases, respectively.

The greatest improvement in interrater reliability with the training module was in the scoring of LMD. Before training, raters were least consistent when labeling LMD ($K = 0.33$) compared with DR ($K = 0.47$) and LR ($K = 0.45$). The training module increased agreement for all recurrence subgroups (Table 2). However, the greatest improvement was for diagnosis of LMD (+67%). LMD IRR more than doubled among non-CNS specialists (+138%). Subgroup analysis suggests that much of the LMD improvement was driven by better recognition of nLMD ($K = 0.60$) versus cLMD ($K = 0.40$), although both were improved relative to the pretraining period. Posttraining agreement was similarly strong for LR ($K = 0.55$) and DR ($K = 0.58$).

After the training module, distant brain relapse was more likely to be scored as nLMD. Rater agreement increased after training as a result of raters changing a portion of their responses. Twenty-seven percent of cases originally labeled as DR were changed to nLMD after the training (Table 2). Similarly, 29% of LR labels were reassigned to nLMD after training. In contrast, very few cases originally labeled as LMD were subsequently relabeled to LR (2%) or DR (6%).

Confidence in assigned labels

To determine which cases raters found challenging, we assessed patterns in the confidence scores raters provided (on a 0–4 scale) for each case. There was a significant relationship between the pretraining assigned pattern of recurrence and confidence (F[2, 657] = 6.81, P = .001). Posthoc comparisons using the Tukey HSD test indicated that the mean confidence for LR (M = 2.5, SD = 1.3) was significantly lower than for LMD (M = 2.9, SD = 1.0, P = .001) and DR (M = 2.9, SD = 1.1, P = .008). Confidence levels between LMD and DR labels did not differ.

There was also a significant relationship between the posttraining assigned pattern of recurrence and confidence (F[3596] = 4.188, P = .003). Mean confidence for cLMD (M = 2.7, SD = 1.1) was significantly lower than mean confidence for DR (M = 3.1, SD = 0.9, P = .007) and nLMD (M = 3.1, SD = 0.9, P = .004). All other comparisons, including those involving LR (M = 2.8, SD = 1.0), were nonsignificant.

At the individual case-rater level, global case-level confidence significantly increased (4.9%, P = 8.04×10^{-5}). Confidence significantly increased for cases originally labeled as LR (7.7%, P = .039) and LMD (7.4%, P < .001). The increases were not significant for DR (1.3%, P = .223).

Discussion

As highlighted by the Response Assessment in Neuro-Oncology working group on LMD, a lack of standardization exists in determining response to treatment.[24,36] Before one can assess response of LMD, one must be able to reliably identify LMD. Our results demonstrate the wide range of physicians' interpretation of neuro-imaging when evaluating intracranial recurrence in patients who received cavity SRS. In our initial assessment, rater agreement was weakest in cases involving LMD compared with either LR or DR. This uncertainty likely contributes to the wide range of LMD risks reported in the literature.[9,10,13,18,19] Our results also show that an approximately 5-minute self-guided training module on identifying nLMD was effective both at increasing interrater agreement across all patterns of recurrence and at increasing raters' confidence in the accuracy of their neuroimaging assessment.

We originally hypothesized that diversity in the literature was driven in part by underrecognition of nLMD in particular. Our results support this hypothesis and in Table 2 we show that 27% of all cases that were originally marked as DR were later revised to nLMD. Altogether, 41% of the cases that would ultimately be categorized as nLMD were originally mislabeled (Table E1, available at https://doi.org/10.1016/j.ijrobp.2019.10.002). Although agreement improved across all patterns of recurrence, the mislabeling rates for LR (15%) and DR (11%) were not as dramatic, suggesting that LMD uncertainty, particularly the nodular subset, drove much of the pretraining discordance.

It is perhaps unsurprising that physicians who specialized in treating CNS-related disease were more consistent as a group compared with non-CNS specialists. Despite the observed label reassignments, for all but one case (LR vs nLMD) the group-wide consensus label was

identical before and after training. Thus, although agreement was weaker before training, the majority of CNS physicians consistently recognized nLMD as a form of LMD. The improved IRR after training likely reflects a tightening of consensus. In contrast, non-CNS specialists' group-wide consensus labels changed for many cases after training, likely reflecting an actual shift of consensus. Despite this considerable reassignment and improved reliability, non-CNS specialists' group-wide assessments for multiple cases differed from both CNS physicians and the unblinded assessment of the senior author—even after training. CNS physicians' group-wide assessments were identical to the unblinded senior author assessments for all but one case (cLMD vs nLMD). Overall our data suggest that when labeling patterns of recurrence is part of a prospective trial, centralized image review by CNS-specific physicians would be preferable to noncentralized review.

The recent LMD guidelines by the Response Assessment in Neuro-Oncology group have described that LMD may sometimes be nodular in appearance.[24] Other work has also shown that cranial or spinal LMD may be nodular.[22,37] However, many sources continue to reference only the diffuse pattern of LMD, which we have accordingly called *classical* LMD. Crucially, even those sources that acknowledge the nodular variant of LMD fail to also provide guidance on distinguishing this morphology from parenchymal metastases—a common source of confusion seen in the present study. This inconsistency, combined with the results from our study, supports the view that nLMD is an underrecognized form of LMD. The variance in reported LMD rates among studies may simply reflect clinicians' differing levels of nLMD recognition.

The exact etiology of nLMD versus cLMD is unclear. Previous studies have found increased LMD in patients with cavity SRS whose resection was piecemeal instead of en bloc,[16,38] suggesting that during resection, malignant cells enter the CSF and settle on the meninges, forming nodules. Preresection SRS may reduce rates of LMD by treating the tumor before surgical perturbation.[39]

Further research will improve salvage outcomes and guide management for patients with nLMD versus DR or cLMD. A recent multi-institutional retrospective study found that 75% of patients with LMD had a neurologic death (personal communication),[28] similar to the 72% seen in Cagney et al.[26] Clinical trials are needed to determine how best to minimize this risk of postresection LMD and neurologic death. Using these and other published data to guide clinical care rests on the assumption of consistency between studies and among physicians in recognition of outcomes. The brief self-guided training module described here is an effective guide for distinguishing nLMD and also improving physician agreement across all 4 patterns of intracranial recurrence (see File A—Training Module, available at https://doi.org/10.1016/j.ijrobp.2019.10.002). This training module will be included in the prospective trial Alliance #A071801 "Phase III Trial of Post-Surgical Single Fraction Stereotactic Radiosurgery (SRS) Compared with Fractionated SRS (FSRS) for Resected Metastatic Brain Disease," where the patterns of failure are scored by local investigators.

Our proposed classification will benefit from further validation and refinement as new data are reported. Given the limited data available at present, we suggest that future retrospective and prospective analyses of brain metastases report all 4 patterns of intracranial recurrence:

(1) LR; (2) DR; (3) nLMD; and (4) cLMD (see Table 1), with consideration to further classify both nLMD and cLMD as localized versus diffuse. Additionally, to best learn how to treat those with the various types of LMD, data on outcomes of salvage treatment are needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Ostrom QT, Wright CH, Barnholtz-Sloan JS. Brain metastases: Epidemiology. Handb Clin Neurol 2018;149:27–42. [PubMed: 29307358]

2. Armstrong JG, Wronski M, Galicich J, Arbit E, Leibel SA, Burt M. Postoperative radiation for lung cancer metastatic to the brain. J Clin Oncol 1994;12:2340–2344. [PubMed: 7964950]

3. Aoyama H, Shirato H, Tago M, et al. Stereotactic radiosurgery plus whole-brain radiation therapy vs stereotactic radiosurgery alone for treatment of brain metastases. JAMA 2006;295:2483. [PubMed: 16757720]

4. Soltys SG, Adler JR, Lipani JD, et al. Stereotactic radiosurgery of the postoperative resection cavity for brain metastases. Int J Radiat Oncol 2008;70:187–193.

5. Iwai Y, Yamanaka K, Yasui T. Boost radiosurgery for treatment of brain metastases after surgical resections. Surg Neurol 2008;69:181–186. [PubMed: 18261647]

6. Mathieu D, Kondziolka D, Flickinger JC, et al. Tumor bed radiosurgery after resection of cerebral metastases. Neurosurgery 2008;62: 817–824. [PubMed: 18414136]

7. Brennan C, Yang TJ, Hilden P, et al. A phase 2 trial of stereotactic radiosurgery boost after surgical resection for brain metastases. Int J Radiat Oncol 2014;88:130–136.

8. Brown PD, Jaeckle K, Ballman KV, et al. Effect of radiosurgery alone vs radiosurgery with whole brain radiation therapy on cognitive function in patients with 1 to 3 brain metastases. JAMA 2016;316:401. [PubMed: 27458945]

9. Brown PD, Ballman KV, Cerhan JH, et al. Postoperative stereotactic radiosurgery compared with whole brain radiotherapy for resected metastatic brain disease (NCCTG N107C / CEC • 3): A multicentre, randomised, controlled, phase 3 trial. Lancet Oncol 2017;18:1049–1060. [PubMed: 28687377]

10. Atalar B, Modlin LA, Choi CYH, et al. Risk of leptomeningeal disease in patients treated with stereotactic radiosurgery targeting the postoperative resection cavity for brain metastases. Int J Radiat Oncol Biol Phys 2013;87:713–718. [PubMed: 24054875]

11. Lamba N, Muskens IS, DiRisio AC, et al. Stereotactic radiosurgery versus whole-brain radiotherapy after intracranial metastasis resection: A systematic review and meta-analysis. Radiat Oncol 2017;12.

12. Patel KR, Prabhu RS, Kandula S, et al. Intracranial control and radiographic changes with adjuvant radiation therapy for resected brain metastases: Whole brain radiotherapy versus stereotactic radiosurgery alone. J Neurooncol 2014;120:657–663. [PubMed: 25189789]

13. Foreman PM, Jackson BE, Singh KP, et al. Postoperative radiosurgery for the treatment of metastatic brain tumor: Evaluation of local failure and leptomeningeal disease. J Clin Neurosci 2018;49:48–55. [PubMed: 29248376]

14. Huang AJ, Huang KE, Page BR, et al. Risk factors for leptomeningeal carcinomatosis in patients with brain metastases who have previously undergone stereotactic radiosurgery. J Neurooncol 2014;120:163–169. [PubMed: 25048529]

15. Jo K-I, Lim D-H, Kim S-T, et al. Leptomeningeal seeding in patients with brain metastases treated by gamma knife radiosurgery. J Neurooncol 2012;109:293–299. [PubMed: 22610938]

16. Suki D, Hatiboglu MA, Patel AJ, et al. Comparative risk of leptomeningeal dissemination of cancer after surgery or stereotactic radiosurgery for a single supratentorial solid tumor metastasis. Neurosurgery 2009;64:664–676. [PubMed: 19197219]

17. Siomin VE, Vogelbaum MA, Kanner AA, Lee S-Y, Suh JH, Barnett GH. Posterior fossa metastases: Risk of leptomeningeal disease when treated with stereotactic radiosurgery compared to surgery. J Neurooncol 2004;67:115–121. [PubMed: 15072456]

18. Mahajan A, Ahmed S, McAleer MF, et al. Post-operative stereotactic radiosurgery versus observation for completely resected brain metastases: A single-centre, randomised, controlled, phase 3 trial. Lancet Oncol 2017;18:1040–1048. [PubMed: 28687375]

19. Soltys SG, Seiger K, Modlin LA, et al. A phase I/II dose-escalation trial of 3-fraction stereotactic radiosurgery (SRS) for large resection cavities of brain metastases. Int J Radiat Oncol 2015;93:S38.

20. Singh SK, Leeds NE, Ginsberg LE. MR imaging of leptomeningeal metastases: comparison of three sequences. AJNR Am J Neuroradiol 2002;23:817–821. [PubMed: 12006284]

21. Chamberlain MC. Comparative spine imaging in leptomeningeal metastases. J Neurooncol 1995;23:233–238. [PubMed: 7673985]

22. Shah LM, Salzman KL. Imaging of spinal metastatic disease. Int J Surg Oncol 2011;1–12.

23. Arora A, Puri S, Upreti L. Leptomeninenal carcinomatosis. In: Brain imaging: Case review series. New Delhi, India: Jaypee Brothers Medical Pub; 2011. p. 122.

24. Chamberlain M, Junck L, Brandsma D, et al. Leptomeningeal metastases: A RANO proposal for response criteria. Neuro Oncol 2017; 19:484–492. [PubMed: 28039364]

25. Freilich RJ, Krol G, Deangelis LM. Neuroimaging and cerebrospinal fluid cytology in the diagnosis of leptomeningeal metastasis. Ann Neurol 1995;38:51–57. [PubMed: 7611725]

26. Cagney DN, Lamba N, Sinha S, et al. Association of neurosurgical resection with development of pachymeningeal seeding in patients with brain metastases. JAMA Oncol 2019;5:703. [PubMed: 30844036]

27. Sneed PK, Mendez J, Vemer-van den Hoek JGM, et al. Adverse radiation effect after stereotactic radiosurgery for brain metastases: Incidence, time course, and risk factors. J Neurosurg 2015;123:373–386. [PubMed: 25978710]

28. Prabhu RS, Turner BE, Asher AL, et al. A multi-institutional analysis of presentation and outcomes for leptomeningeal disease recurrence after surgical resection and radiosurgery for brain metastases. Neuro Oncol 2019;21:1049–1059. [PubMed: 30828727]

29. Le Rhun E, Weller M, Brandsma D, et al. EANO—ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up of patients with leptomeningeal metastasis from solid tumours†. Ann Oncol 2017; 28:iv84–iv99. [PubMed: 28881917]

30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377–381. [PubMed: 18929686]

31. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378–382.

32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159. [PubMed: 843571]

33. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. Phys Ther 2005;85:257–268. [PubMed: 15733050]

34. McHugh ML. Interrater reliability: the kappa statistic. Biochem medica 2012;22:276–282.

35. Tukey JW. Comparing individual means in the analysis of variance. Biometrics 1949;5:99. [PubMed: 18151955]

36. Chamberlain M, Soffietti R, Raizer J, et al. Leptomeningeal metastasis: A response assessment in neuro-oncology critical review of endpoints and response criteria of published randomized clinical trials. Neuro Oncol 2014;16:1176–1185. [PubMed: 24867803]

37. Smirniotopoulos JG, Murphy FM, Rushing EJ, Rees JH, Schroeder JW. Patterns of contrast enhancement in the brain and meninges. Radiographics 2007;27:525–551. [PubMed: 17374867]

38. Suki D, Abouassi H, Patel AJ, Sawaya R, Weinberg JS, Groves MD. Comparative risk of leptomeningeal disease after resection or stereotactic radiosurgery for solid tumor metastasis to the posterior fossa. J Neurosurg 2008;108:248–257. [PubMed: 18240919]

39. Patel KR, Burri SH, Asher AL, et al. Comparing preoperative with postoperative stereotactic radiosurgery for resectable brain metastases. Neurosurgery 2016;79:279–285. [PubMed: 26528673]
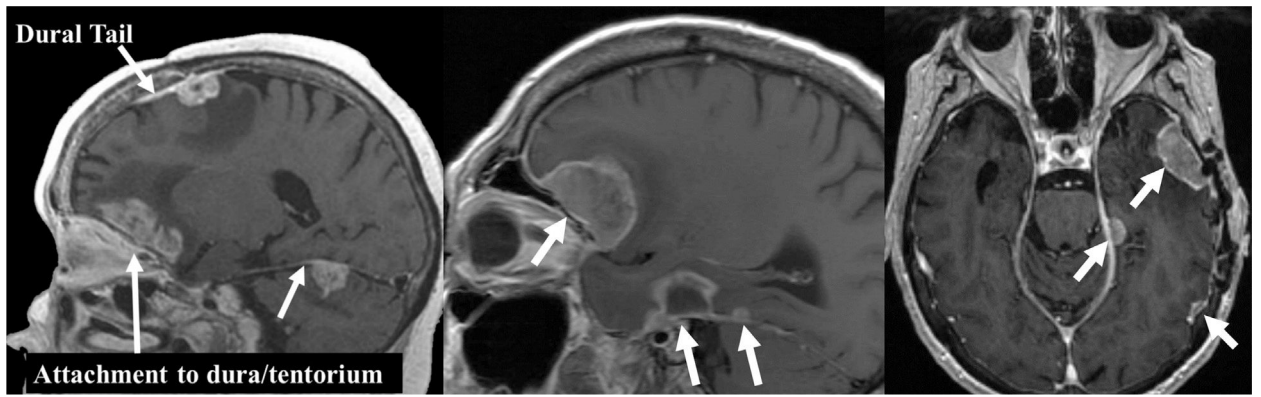
Fig. 1.
Sample image from the training module providing guidance on how to differentiate different patterns of recurrence for the study. This is one of the images highlighting common features in patients with nodular leptomeningeal disease.
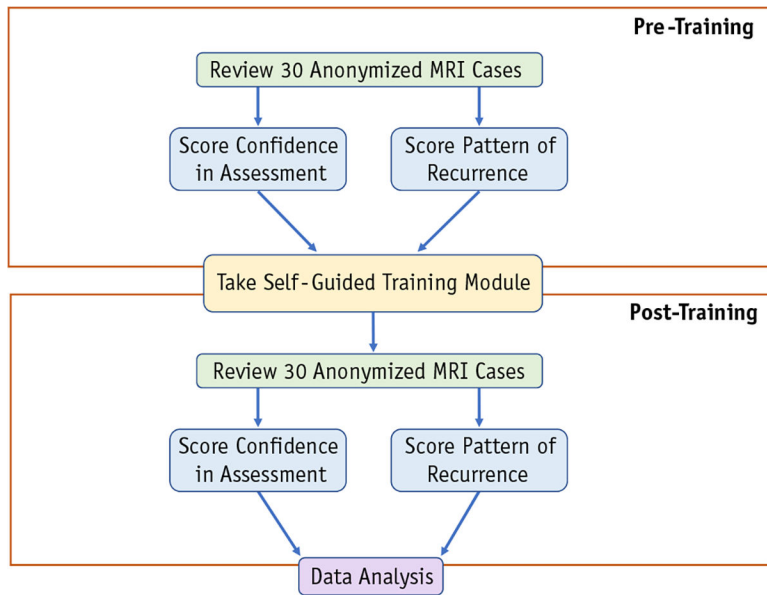
Fig. 2.
Schematic overview of data generation workflow for each physician rater. The 30 anonymized magnetic resonance imaging (MRI) cases were randomized before each review. Pattern of recurrence labeling was multiple choice. Pretraining recurrence options: local recurrence, distant parenchymal recurrence, and leptomeningeal disease. Posttraining recurrence options: local recurrence, distant parenchymal recurrence, nodular leptomeningeal disease, and classical leptomeningeal disease. Raters assigned a confidence score to each answer choice on a scale of 0 to 4.

Table 1

Proposed classification of intracranial progression after treatment for brain metastases

Intracranial disease	Supportive neuroimaging features[*]
Classical leptomeningeal disease (cLMD)	Enhancement of cranial nerves Curvilinear enhancement within: Cerebellar folia Cerebral sulci Cerebral cisterns Diffuse "sugar coating" of the surface of the brain
Nodular leptomeningeal disease (nLMD)	Focal nodule(s) adherent to surfaces with CSF contact: Dural/pial surface Tentorium Ventricles Hypervascular dural tail
Distant intraparenchymal metastases	Focal lesion deep to the pial surface Hematogenous spread pattern
Local recurrence	Nodular enhancement within the resection cavity In-field recurrence within the 80% isodose line

Note: cLMD and nLMD can be further subjectively described as "localized" or "disseminated."

[*] This column summarizes the pattern of recurrence radiographic labeling guidelines provided to raters in the training modules.

Table 2

Rater agreement summary Statistics

	All raters	CNS specialists	Non-CNS specialists
Total	22	15	7
Pretraining IRR			
Overall	0.41	0.46	0.35
LR	0.45	0.49	0.36
DR	0.47	0.52	0.46
LMD	0.33	0.39	0.21
Posttraining IRR			
Overall	0.55	0.56	0.56
LR	0.55	0.59	0.52
DR	0.58	0.61	0.52
cLMD	0.40	0.47	0.37
nLMD	0.60	0.56	0.73
IRR after training			
Overall	+34% [*]	+22% [*]	+60% [*]
LR	+22% [*]	+20% [*]	+44% [*]
DR	+23% [*]	+17% [†]	+13%
LMD	+67% [*]	+44% [‡]	+138% [‡]
Reassignment			
DR —> nLMD	27%	30%	28%
LR —> nLMD	29%	27%	27%
LMD —> DR	6%	6%	6%
LMD —> LR	2%	1%	6%

Abbreviations: cLMD = classic leptomeningeal disease; CNS = central nervous system; DR = distal recurrence; IRR = interrater reliability; LMD = leptomeningeal disease; LR = local recurrence; nLMD = nodular leptomeningeal disease.

Summary statistics for IRR and case label reassignment results. Before training, IRR was lowest for LMD, particularly among non-CNS specialists (0.21). The training module improved IRR by 34% for all raters ($P<.001$), with an even greater improvement among non-CNS specialists (IRR + 60%, $P< .001$) compared with CNS specialists (IRR + 22%, $P< .001$). Subgroup change analyses (LR, DR, LMD) used a Bonferroni correction to account for multiple comparisons. IRR increased across all subgroups, with the greatest benefit seen in LMD cases. The improved IRR was driven largely by improved recognition and reassignment of cases with nLMD. IRR was assessed using Fleiss' Kappa statistic. Significance of the pairwise overall IRR percentage change pre- and posttraining was assessed using Wilcoxon's signed rank test. Significance of subgroup IRR change was not assessed due to lack of pairwise comparisons.

[*] $P< .001$.

[†] $P< .05$.

[‡] $P< .01$.