Neuro-Oncology Advances
4(1), 1–11, 2022 | https://doi.org/10.1093/noajnl/vdac116 | Advance Access date 26 July 2022

Real-time PACS-integrated longitudinal brain metastasis tracking tool provides comprehensive assessment of treatment response to radiosurgery

Gabriel Cassinelli Petersen, Khaled Bousabarah, Tej Verma, Marc von Reppert, Leon Jekel, Ayyuce Gordem, Benjamin Jang, Sara Merkaj, Sandra Abi Fadel, Randy Owens, Antonio Omuro, Veronica Chiang, Ichiro Ikuta, MingDe Lin, and Mariam S. Aboian

Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, Connecticut, USA (G.C.P., M.v.R., L.J., A.G., B.J., S.M., S.A.F., I.I., M.L., M.S.A.); University of Göttingen Medical Faculty, Göttingen, Germany (G.C.P.); Visage Imaging GmbH, Berlin Germany (K.B.); New York University, New York City, New York, USA (T.V.); Visage Imaging Inc., San Diego, California, USA (R.O., M.L.); Department of Neurology, Yale School of Medicine, New Haven, Connecticut, USA (A.O.); Department of Neurosurgery, Yale School of Medicine, New Haven, Connecticut, USA (V.C.); Yale Program for Innovation in Imaging Informatics, Yale School of Medicine, New Haven, Connecticut, USA (M.S.A., I.I.)

Corresponding Author: Mariam S. Aboian, MD/PhD, Department of Radiology, Yale School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA (mariam.aboian@yale.edu)

Abstract

Background: Treatment of brain metastases can be tailored to individual lesions with treatments such as stereotactic radiosurgery. Accurate surveillance of lesions is a prerequisite but challenging in patients with multiple lesions and prior imaging studies, in a process that is laborious and time consuming. We aimed to longitudinally track several lesions using a PACS-integrated lesion tracking tool (LTT) to evaluate the efficiency of a PACS-integrated lesion tracking workflow, and characterize the prevalence of heterogenous response (HeR) to treatment after Gamma Knife (GK).

Methods: We selected a group of brain metastases patients treated with GK at our institution. We used a PACS-integrated LTT to track the treatment response of each lesion after first GK intervention to maximally seven diagnostic follow-up scans. We evaluated the efficiency of this tool by comparing the number of clicks necessary to complete this task with and without the tool and examined the prevalence of HeR in treatment.

Results: A cohort of eighty patients was selected and 494 lesions were measured and tracked longitudinally for a mean follow-up time of 374 days after first GK. Use of LTT significantly decreased number of necessary clicks. 81.7% of patients had HeR to treatment at the end of follow-up. The prevalence increased with increasing number of lesions.

Conclusions: Lesions in a single patient often differ in their response to treatment, highlighting the importance of individual lesion size assessments for further treatment planning. PACS-integrated lesion tracking enables efficient lesion surveillance workflow and specific and objective result reports to treating clinicians.

Key points

1. Our PACS-integrated lesion tracking tool allows efficient tracking of multiple lesions over long time periods. It also allows specific, and objective reporting of results.
2. Lesions in a single patient often respond to treatment heterogenously.

Importance of the Study

This study introduces a workflow efficient PACS-integrated lesion tracking tool to the neuro-oncological community. This tool substantially facilitates the neuroradiologist's task of surveilling lesion growth, a process that is particularly challenging in patients with multiple simultaneous lesions that are being monitored over multiple follow-up studies. Through generation of treatment response curves and tables, the neuroradiologist is also able to communicate changes in lesion sizes in a concise, objective, and specific manner to clinicians planning individualized treatment. Furthermore, we also utilized the efficient workflow enabled by the tool to track close to 500 lesions over long periods of time and characterize the treatment response to stereotactic radiosurgery. We found that lesions in a single patient often respond to treatment heterogenously, a finding that underscores the necessity for assessing treatment response at individual lesion level.

Brain metastases comprise more than 50% of intracranial tumors in adults.[1] Targeted management of metastatic lesions in the brain has become available in recent decades thanks to the development of stereotactic radiosurgical techniques such as Gamma Knife (GK). However, correctly assessing which lesions should be targeted often depends on the radiologist correctly tracking the change in size of individual lesions over time and communicating these results to the treating clinician in a clear and efficient manner. This process is especially challenging in the setting of brain metastases because patients frequently present with multiple lesions over a number of prior MR scans. This often results in long reading times and the radiologists losing track of lesions, all of which are factors that might be contributing to the increasing burn-out rate among neuroradiologists.[2] These also contribute to the generation of unclear and nonspecific radiological reports, which are at odds with the need of clinicians to precisely define the behavior of individual lesions for treatment planning purposes.[3] In addition, this shifts the responsibility of tumor treatment response assessment to neuro-oncologists, who may not have ready access to measurement tools needed to generate treatment response curves of tumors. Furthermore, objective and efficient communication between radiologists and clinicians is hampered by the free-text method of report writing, the current standard for reporting treatment responses. As pointed out by Bink and colleagues, the use of this method is especially problematic in follow-up readings of intracranial pathologies—such as brain metastases—since the inter-reader variability is high and the reproducibility of diagnostic measures is hindered.[4] Considering these problems, we sought to develop an easy-to-use and efficient PACS-integrated method for longitudinal lesion tracking that enables radiologists to systematically track the growth of multiple individual lesions over a long period of time, and facilitates communication of results to referring clinicians: both qualitatively through automatic generation of treatment response curves, and quantitatively by summarizing longitudinal changes in tabular form. We also aimed to use the lesion tracking tool (LTT) to determine the prevalence of heterogenous response (HeR), meaning the co-occurrence of different response categories in lesions of the same patient treated with GK radiosurgery.[5]

Materials and Methods

Patient Selection

This was a retrospective (informed consent was waived), IRB approved, HIPAA compliant single-center study. To characterize the performance of the LTT and radiosurgery treatment response assessment vs standard of care (ie manual measurements, response assessment, and reporting), we identified a test dataset of patients with brain metastases. The selection process workflow is summarized in Figure 1. Briefly, we obtained the patient list from Brain Metastases Tumor Board meetings at our institution from August 2020 to June 2021, and randomly selected a group of 100 patients with brain metastases who had radiosurgery at the institution. Further inclusion criteria were (i) an initial GK scan, and (ii) a minimum of two follow-up diagnostic MR scans at least one month apart from each other. Selected exclusion criteria were (i) postcontrast T1-weighted (T1c+) images with a slice thickness ≥3mm, and (ii) patients that only had resection cavities due to pre-GK surgical excision of metastases, because it is known that rim-enhancing resection cavities are difficult to measure in 2D, and the cavity size reflects the extent of surgery, and less the disease itself.

PACS-Integrated LTT

Studies that matched the inclusion criteria were transferred using the DICOM send function from the clinical production PACS to a research instance of the PACS (AI Accelerator, Visage Imaging, Inc., San Diego, CA). The images were de-identified using a nonreversible hash-based process whereby if another study for the same patient is sent through the pipeline later, those new objects are assigned to the same de-identified patient jacket on the research server, facilitating lesion tracking of de-identified data. The LTT that is included in Visage7 (Visage Imaging Inc., San Diego, CA) was used for our study. 1D and 2D lesion diameters can be measured, and the results are stored as DICOM Structured Report in PACS. Furthermore, each lesion is tracked using a unique identifier so that all prior measurements can be automatically loaded upon opening a patient's current study. The lesion measurements across

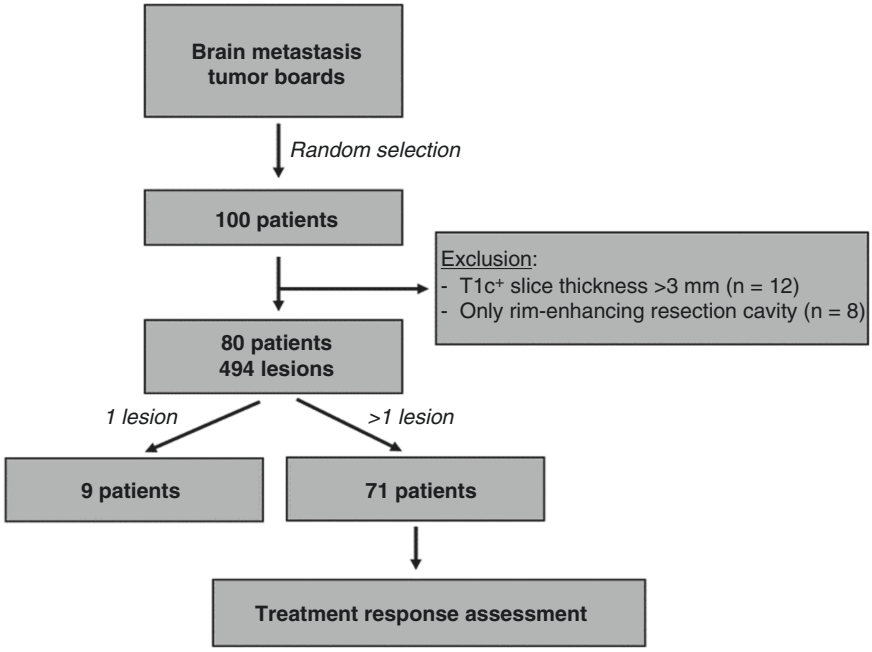


Figure 1. Patient selection workflow. From the randomly selected 100 patients, 80 were included in the analysis and measured longitudinally. From this group, nine had only one lesion and homogeneity of treatment response could be assessed in seventy-one patients.

time can then be automatically quantified and visualized using both tables and growth curves. To accelerate the workflow, we implemented a custom hanging protocol that can automatically co-register, chronologically align, and present an eight-viewer layout that includes the current and up to seven prior image series acquired using T1c+ gradient-echo sequences.

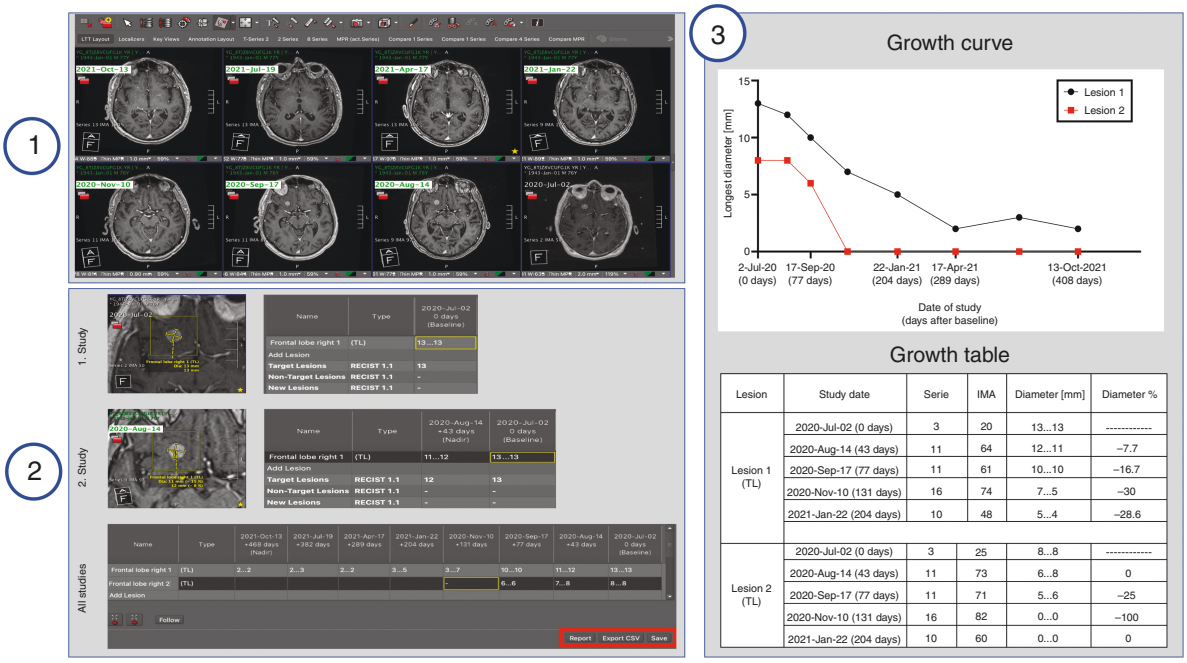
Tracking of Lesions

Measurements and lesion tracking were performed using the LTT on the research PACS. All intraparenchymal metastatic lesions were first measured in the first GK intervention scan and tracked for up to seven follow-up MR scans, unless the patient died before, or the patient had not yet been followed up seven times. Extraparenchymal lesions were excluded because leptomeningeal disease is considered nonmeasurable by the RANO-BM proposal.[6] The process of PACS-integrated longitudinal lesion tracking using the LTT is outlined graphically in Figure 2 and an illustrated step-by-step instruction on its use is presented in Supplementary Result 1. Briefly, after the first study is opened, the user deploys the lesion tracking hanging protocol by the click of a push-button automatically loading all the relevant studies to be measured, automatically aligning the desired sequences chronologically, co-registering them, and opening the LTT interface. This interface is composed of a table with the study dates as column-headings, and individual lesions as row-headings. Next, the 2D measurement tool embedded in the LTT is used to measure lesions. The size is annotated automatically in the corresponding cell of the table. By consecutively measuring the size of the same lesion on different study dates, the progression of a single lesion is annotated in its corresponding row. In addition, the lesion can be renamed and further characterized as a "Target Lesion," "Non-Target Lesion," and "New Lesion." Once all lesions have been measured the user can automatically create and export the output of the LTT as (i) a graphical growth curve of all lesions, where the x-axis represents time and the y-axis the longest diameter (LD) in mm; and (ii) a growth table detailing the name of the lesion, study dates, series, and image number where the lesion was found on a particular study, diameters in mm, and the percentage diameter change in relationship to the previous study. The output can then be pasted into a reporting tool (ie PowerScribe, Nuance, Burlington, MA), or the growth table can be exported as a .csv file. Finally, the results can be saved on PACS in the patient jacket and retrieved at a later date to add the results of the newest follow-up study.

We calculated the total number of patients, lesions, number of lesions per patient, time of follow-up after first GK scan, and number of GK interventions per patient. We summarized the results separately for patients who died within the first seven follow-up scans after first GK intervention, and those who remained alive. We compared these characteristics between the two groups with a two-tailed Mann-Whitney *U* test ($\alpha = 0.05$).

Measuring Efficiency Gains of LTT

To objectively measure the efficiency of the LTT from the viewpoint of a radiologist, we compared the number of mouse clicks needed to measure and track lesions with and



Lesion	Study date	Serie	IMA	Diameter [mm]	Diameter %
Lesion 1 (TL)	2020-Jul-02 (0 days)	3	20	13...13	----------
	2020-Aug-14 (43 days)	11	64	12...11	–7.7
	2020-Sep-17 (77 days)	11	61	10...10	–16.7
	2020-Nov-10 (131 days)	16	74	7...5	–30
	2021-Jan-22 (204 days)	10	48	5...4	–28.6
Lesion 2 (TL)	2020-Jul-02 (0 days)	3	25	8...8	----------
	2020-Aug-14 (43 days)	11	73	6...8	0
	2020-Sep-17 (77 days)	11	71	5...6	–25
	2020-Nov-10 (131 days)	16	82	0...0	–100
	2021-Jan-22 (204 days)	10	60	0...0	0

Figure 2. Workflow of the lesion tracking tool (LTT) for longitudinal tracking of metastases. **Step 1:** Loading of the images and deployment of the hanging protocol. **Step 2:** Individual lesions are measured over time and the results recorded and automatically assigned to their corresponding study date. **Step 3:** The output of the LTT is generated. The automatically generated treatment response curves and tables allow for qualitative and quantitative assessment of lesion growth. In this example, two metastatic lesions are tracked for 408 days after Gamma Knife (GK) radiosurgery over seven follow-ups post GK, revealing that both lesions remitted homogenously after treatment.

without use of the LTT, first in theory and secondly in practice. We first recorded the amount of clicks necessary for the initial process of loading, aligning, and co-registering the images and the measurements itself if a radiologist were to work as efficiently as possible. To recreate a scenario resembling every day clinical practice we counted the number of clicks that two readers needed to track five patients, each with every eight studies (including the first GK scan) and two lesions with and without the LTT. During tracking by a reader, the other counted the number of clicks. These were recorded separately for the process of loading the images (starting with the opening of the first study and ending right before the first lesion size measurement begins) and for the measurements (starting with the selection of the measurement tool and ending when the last lesion has been measured). Significance of difference in number of mouse clicks with and without using the LTT was determined using a paired *t*-test ($\alpha = 0.05$).

Comparison of Radiological Free-Text Impression versus Automatic LTT Output

We randomly selected 30 patients from our cohort and collected the output of the LTT (growth curves and tables), and the original radiological impression at the last date of follow-up. A board-certified neuroradiologist (5 years of experience) was tasked to review both sets of information and answer the following two questions per case: "Regarding the growth of metastatic lesions, are the radiological impressions congruent with the growth curves/tables" (Possible answers: Congruent/Partially Congruent/Incongruent), and "For this case, how much more information relevant to treatment planning does the growth curves/table add to the radiological impression?" (Possible answers: Scale from 1 to 5, where 1 = "no additional relevant information for treatment planning", and 5 = "maximal addition of relevant information for treatment planning"). In cases where the report and LTT output were incongruent, the measurements with the LTT were double-checked to ensure that the discrepancy was not due to a failure in correctly measuring lesions.

Assessing Prevalence of HeR to Treatment at Individual Patient Level

To characterize how brain metastases behave in relation to one another after radiosurgery at an individual patient level, we selected patients that had >1 lesion during the measured timeframe. We aimed to characterize the response as either homogenous (all lesions either increased, decreased, or remained stable after treatment) or as heterogenous (the treatment response of lesions differed within a patient after radiosurgery). To objectify this, all lesions were classified at every available follow-up scan as either stable disease, partial response (PR), complete response, progressive disease, or as new lesions (NL) if they appeared

Table 1. Characteristics of patients included in this study.

	All patients	Alive patients	Dead patients
Total number of patients	80	49	31
Mean age (± SD)	63 years(±13)	63.6 years (±12.9)	62.7 years(±13.6) **[*P* = .9, ns]**
F:M ratio	41:39	24:25	17:14
Total number of lesions	494	237	257
Median number of lesions per patient (range)	4 lesions/patient(1-45)	3 lesions/patient(1-26)	4 lesions/patient(1-45) **[*P* = .101, ns]**
Mean time of follow-up after first GK (± SD)	373.6 days(±176.1)	457.2 days(±176.1)	241.5 days(±145.6) **[*P* < .001, significant]**
Median number of GK interventions per patient (range)	1 intervention/patient(1-4)	1 intervention/patient(1-3)	1 intervention/patient(1-4) **[*P* = .353, ns]**
Primary cancer	NSCLC = 37 SCLC = 4 Melanoma = 17 Renal-CA = 6 Breast-CA = 6 Prostate-CA = 2 GIT-CA = 6 Oropharyngeal-CA = 1 Ovarian-CA = 1	NSCLC = 20 SCLC = 2 Melanoma = 14 Renal-CA = 5 Breast-CA = 4 Prostate-CA = 1 GIT-CA = 3	NSCLC = 17 SCLC = 2 Melanoma = 3 Renal-CA = 1 Breast-CA = 2 Prostate-CA = 1 GIT-CA = 3 Oropharyngeal-CA = 1 Ovarian-CA = 1

The column titled "All patients" summarizes the information among all 80 patients that were included in this study. The column "Alive patients" includes all patients that remained alive during the entirety of the follow-up of up to seven follow-up scans after first GK intervention. The column "Dead patients" summarizes all patients who died during follow-up.
Abbreviations: CA, carcinoma; F:M ratio, Female-to-Male ratio; GIT, gastrointestinal tract; GK, Gamma Knife; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer.

for the first time. Analogous to the proposed RANO-BM criteria, PR was defined as ≥30% decrease in LD from baseline in lesions with a LD ≥10 mm, and an absolute decrease of ≥3 mm, for lesions with LD <10 mm.[6] Progression was defined as an increase ≥20% in LD from nadir for lesions with an LD ≥10 mm, and an absolute increase ≥3 mm for lesions with a LD <10 mm. When neither criterion applied, lesions were classified as stable.[6] Whenever a NL appeared we annotated it as "New Lesion," recorded the size, and for the following scans classified it as decreasing, increasing, or stable analogously to all other lesions. If all lesions were either partially or completely remitted in a single follow-up study, the treatment response was classified as homogenously decreasing (HD). If all lesions progressed or new ones appeared, treatment response was classified as homogenously increasing (HI). If all lesions remained stable in size, treatment response was classified as homogenously stable (HS). Treatment response was classified as HeR when a combination of decreased and/or increased and/or stable lesions was found on the same follow-up scan. Furthermore, we specified whether this was due to a combination of "stable and increasing lesions," "stable and decreasing lesions," "decreasing and increasing lesions," or "stable, increasing, and decreasing lesions." We assessed the prevalence of the different treatment responses at the last follow-up scan and related this to the number of metastatic lesions the patient had; and at 0–90, 91–180, 181–270, 271–365, >365 days after the first GK intervention. We also performed a secondary analysis in which we excluded from the analysis all lesions with a LD <5 mm because they are considered nonmeasurable by the proposed RANO-BM criteria.[6]

Results

Patient Selection

A tabular summary of the patient characteristics can be found in Table 1. From the initial group of 100 randomly selected patients, 12 were excluded because the T1c+ images had slice thickness ≥3 mm, and eight patients due to only having rim-enhancing resection cavities on the first GK intervention. 80 patients with a total of 494 lesions were included in our study (Mean age ± standard deviation = 63 ± 13 years, female-to-male ratio = 41:39). For a secondary analysis of treatment response assessment only in lesions considered measurable by the RANO-BM criteria we excluded 140 lesions because their LD at baseline was <5 mm. This yielded 79 patients with 354 lesions. All included patients had gradient-echo T1c+ follow-up studies with slice thickness <1.5 mm. Of the 80 patients, 31 (38.75%) died before reaching their respective seventh follow-up MR scan date. The median number of lesions/patient was 3 (Range: 1–26) in the group that remained alive during follow-up, and 4 (Range:1–45) in the group that died. The difference did not reach statistical significance (*P* = .101). The mean ± standard deviation follow-up time was 457.2 ± 176.1 days for patients that remained alive, and

241.5 ± 145.6 days for those who died. The mean follow-up time among patients that had seven follow-up scans after the first GK intervention was 474 ± 116 days (Range: 268–721). The median number of GK interventions for both groups was 1 (Range:1–4). Among all 80 patients, nine had only one intraparenchymal lesion, while the rest had multiple. Of the 79 patients included in our secondary analysis, 63 had more than one lesion.

Measurement of Efficiency of LTT

To objectively and reproducibly measure efficacy gained using the LTT we compared the number of clicks needed for the entire process with and without its use. The results have been summarized graphically in Supplementary Figure 1. To load the previous studies and align them, the LTT requires a total of at least four clicks to open one study, and one additional click per additional study (one to open the study navigator, one to select each study, one to load all selected studies, and one to deploy the hanging protocol). Conversely, to align the studies chronologically and co-register them without LTT it takes at least five clicks to open one study and three additional clicks per additional study (one to open the study navigator, one per study to open all available sequences in one study, one per study to scroll through the sequences, one per study to drag and drop the sequence into its corresponding viewer, and one to co-register all aligned studies). To perform the 2D measurements, it requires in both cases, three clicks per lesion (one to activate the 2D ruler, and two to measure each diameter). At this point and in both cases all lesions will have been measured, but the results are ordered logically only when using LTT. Finally, to communicate the results, LTT users can transfer the growth table into the report, whereas radiologists without LTT will manually dictate the results.

To test the efficiency of the LTT in practice we measured the number of clicks needed for loading and measuring brain metastases when two readers tracked five patients with every eight studies (including the first GK scan) and two lesions. The results are shown in Figure 3. The time needed to dictate individual lesions into the radiologic report was not tracked for this analysis.

Comparison of Radiological Impressions and LTT Output

Regarding the growth of lesions, the original radiologist's impression was congruent or partially congruent in 26.6% (8/30) and 33.3% (10/30) of times. 40% (12/30) of radiological impressions were judged to be incongruent with the LTT output. The additional information provided by the LTT output regarding treatment planning was scored with the maximal score of "5" in 93.3% (28/30) of cases, and "1" and "3" in 3.3% (1/30) of cases each.

Prevalence of HeR to Treatment at Individual Patient Level

Examples of treatment response curves showing different homogenous and HeR can be seen in Figure 4. An example case showing tracking of >25 metastases over multiple follow-ups is demonstrated in Supplementary Figure 2. Results are summarized graphically in Figure 5, and tabularly in Supplementary Tables 1–3. 81.7% (n = 58) of patients with >1 lesion (n = 71) showed a HeR at the last date of follow-up, while 12.7% (n = 9) showed a homogenous decrease in lesions size, 4.2 % (n = 3) a HI, and 1.4% (n = 1) HS response to treatment. 32.8% (n = 19) of patients with a HeR had a combination of decreasing and increasing lesions; 31% (n = 18) stable, decreasing, and increasing lesions; 20.7% (n = 12) stable and decreasing lesions; 15.5% (n = 9) stable and increasing lesions (Figure 5A, Supplementary Table 1A). Prevalence of HeR to treatment was lowest in patients with two lesions (47.1%, n = 8/17), and highest in patients with >4 lesions (100%, 29/29). Patients with three and four lesions had a prevalence of HeR in 81.8% (n = 9/11) and 85.7% of cases (n = 12/14), respectively (Figure 5B, Supplementary Table 2A). In patients with follow-up images in the first 90 days post first GK intervention and more than one lesion (n = 58), 69% (n = 40) showed a HeR, 20.7% (n = 12) HS, and 10.3% (n = 6) HD response to treatment. The fraction of HeR increased in follow-ups 91–180 days (80.6%, n = 50/62), 181–270 days (82%, n = 41/50), and 271–365 (83%, n = 34/41) post GK, and decreased in follow-up scans later than 1 year (77.5%, n = 31/40) post first GK. (Figure 5C, Supplementary Table 3A).

In our secondary analysis which excluded lesions with LD <5 mm at baseline, 71.4% (n = 45) of the patients with more than one lesion (n = 63) showed a HeR at last date of follow-up, 19% (n = 12) showed a homogenous decrease in lesions size, 4.8% (n = 3) showed each HI and HS response to treatment. Among the 45 patients with HeR, 42.2% (n = 19) had a combination of decreasing and increasing lesions; 31.1% (n = 14) stable, decreasing, and increasing lesions; 13.3% (n = 6) each stable and decreasing lesions; and stable and increasing lesions. (Supplementary Table 1B). The prevalence of HeR to treatment was lowest in patients with two lesions (47.8%, n = 11/23), and highest in patients with more than four lesions (90%, 18/20). Patients with three and four lesions had HeR in 83.3% (n = 10/12) and 75% of cases (n = 6/8), respectively (Supplementary Table 2B). In patients with follow-up images taken in the first 90 days post first GK intervention and >1 lesion (n = 49), 61.2% (n = 30) showed a HeR, 28.6% (n = 14) HS, and 10.2% (n = 5) HD response to treatment. The fraction of HeR increased in follow-ups 91–180 days (72.2%, n = 39/54), and 181–270 days (75%, n = 33/44) post GK, and decreased in follow-up scan 271–365 (71.4%, n = 25/35),and later than 1 year (63.6%, n = 21/33) post first GK (Supplementary Table 3B).

Discussion

In this study, our group aimed to develop an easy-to-use and efficient method for longitudinal lesion tracking that enables radiologists to systematically track the change in size of multiple individual lesions over a long period of time and facilitates communication of results to referring clinicians. We also leveraged the LTT to determine the prevalence of HeR in lesions treated with radiosurgery.

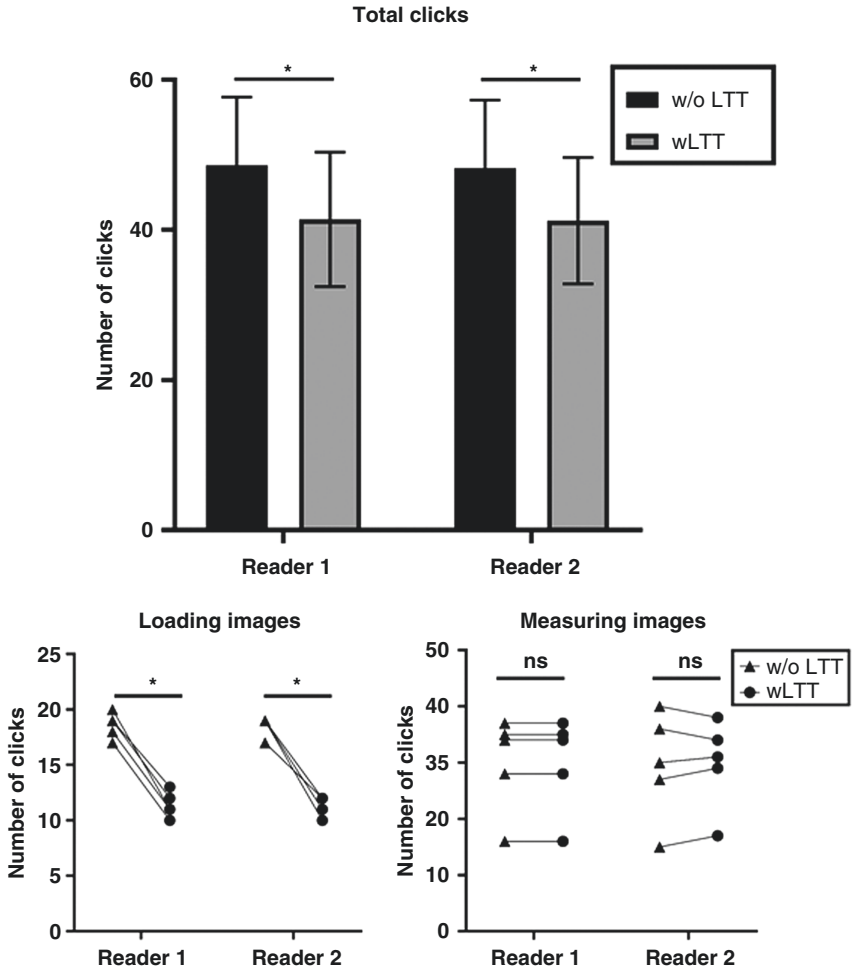


Figure 3. Comparison of clicks necessary to measure and track lesions with and without the LTT. Mean total click count was significantly lower with use of LTT (Reviewer 1: 48.6 vs 41.4 clicks, $P < .001$; Reviewer 2: 48.2 vs 41.2 clicks, $P < .001$). Similarly, mean click count for loading the images was significantly lower when the LTT was used (Reviewer 1 and 2: 11.4 vs 18.6 clicks, $P < .001$). There was no significant difference between the number of clicks necessary for the size measurement of the lesions itself (Reviewer 1: 30 vs 30 clicks; Reviewer 2: 29.8 vs 29.6 clicks, $P = .84$). * = $P < .001$, ns, not significant.

The LTT is fully integrated in PACS. This is necessary to facilitate longitudinal lesion tracking since it allows its seamless integration into the clinical workflow of radiologists. Furthermore, this integration allows the LTT to be easily accessed by deployment of a hanging protocol that (i) identifies the desired sequences from previously loaded studies automatically, (ii) auto-aligns them chronologically, (iii) co-registers them automatically, and (iv) automatically opens the LTT interface. Importantly, while we configured our hanging protocol to retrieveT1c+ gradient-echo images, the protocol can be modified to retrieve other sequences (eg, FLAIR images) and allow longitudinal tracking of different lesions (eg, demyelinating lesions). The efficiency of this hanging protocol is also supported by our finding that it significantly improved the radiologist's workflow by decreasing the amount of clicks necessary to measure lesions longitudinally. To measure the lesions, the radiologist follows the same procedure as to measure any structure without the LTT, but the measurement is automatically recorded by the tool and annotated in the interface table. In other words, the way the radiologist measures the lesions is unchanged but has the added benefit of recording the measurement automatically. Furthermore, thanks to the side-by-side layout of prior studies, visual comparison of lesions is facilitated. In addition, the co-registration of images allows the tool to closely predict the image number at which the lesion is going to be found in future follow-ups, and guide the radiologist, saving time finding the lesion. Importantly, this tool can not only be used retrospectively—as in our study—but also prospectively since the measurements can be saved and updated later when the patient comes for the next follow-up appointment. Upon completion of the measurements, the provider can generate two types of output: a graphical treatment response curve that conveys changes in lesion size qualitatively; and treatment response tables that quantitatively details the size variations. Both types of output can then be added to the radiological report providing clinicians with an objective, specific

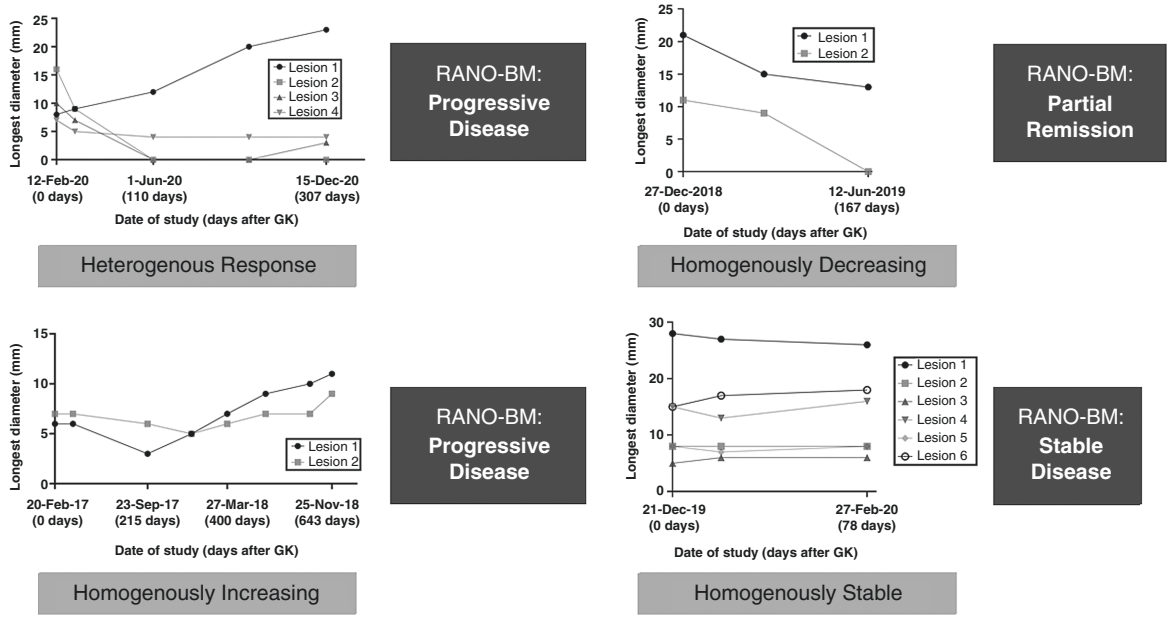


Figure 4. Types of treatment response after GK therapy demonstrating growth curves with homogenous and heterogenous treatment responses in comparison to treatment response calculated by RANO-BM. 3A and 3B demonstrate progressive disease as classified by RANO-BM criteria, but when looking at growth curves, 3A demonstrates heterogenous response with only one lesion needing treatment. On the other hand, 3B demonstrates homogeneously increasing growth curves, thus there is a need to treat both lesions in this patient. 3C demonstrates partial remission with homogeneous decrease in size of all lesions over time. 3D demonstrates stable disease on both RANO-BM and growth curve assessment. These treatment response curves show that while an overall assessment of treatment response is necessary, in some cases it is also advantageous to look at the change in size of individual lesions: while the overall sum of diameters is increasing in Lesion 1 and slightly in Lesion 3, Lesions 2 and 4 have both remitted or are remained stable after treatment.

and consistent structured report, while also saving dictation time. Structured reports have been previously evaluated in the field of neuro-oncology and have been found to reduce nonspecific descriptions of brain lesions in patients with brain metastases to help detect all intracranial pathologies, and reproducibly document findings.[3,4] In our study comparing the free-text radiologist impressions with the structured output provided by the LTT, we found inconsistencies regarding the growth of lesions. In one striking case with multiple metastases, a lesion increased in size from 14 to 19 mm in LD, a change that was clearly shown in the treatment response curve and even recorded by the radiologist in the report but was left out from the impression of the radiology report. Another example demonstrating relevant incongruencies between the original radiologist's impression and the corresponding growth curve is showcased and discussed in Supplementary Figure 3. Furthermore, the additional information gained from the LTT output was overwhelmingly interpreted to be most important for treatment planning purposes. While these results suggest that the LTT could increase the diagnostic accuracy of brain metastases, a controlled clinical trial will be necessary to prove this. LTT can also be applied to other diseases like liver, lung, etc., lesion assessment where a streamlined longitudinal lesion tracking workflow can help alleviate radiologist burnout, make for more consistent measurements and effective reporting, and adoption of ACR (LI-, PI-, TY-, etc.) RAD guidelines. In addition, quantitative output can be used to develop and incorporate novel treatment response criteria such as brain metastasis velocity in clinical workflow.[7]

Overall, the integration of the longitudinal brain metastasis measurement tool into PACS, the development of the hanging protocol, the easy-to-use interface and the possibility to communicate the results in a clear, objective and consistent manner should encourage neuroradiologists to longitudinally track different lesions over multiple priors and allow for these assessments to become standard of care. In a study published by Hayward and colleagues in 2016, the authors were able to show that taking multiple prior mammography studies into account resulted in reduced recall rates, an important benefit for patients.[8] While no study examining the benefit of considering multiple prior MR scans in the context of brain metastases has been published to our knowledge, it is very likely that neuro-oncologists, radiation oncologists, and neurosurgeons will be able to tailor treatment plans more accurately if the behavior of metastases can be examined and interpreted not just in relation to the last scan but over longer time periods. Prior studies have shown that metastatic brain lesions vary in their responses substantially: while the lesions commonly decrease in size, a large proportion remains stable or increases in size.[9,10] In this context, we sought to utilize our newly developed LTT and characterize the different forms of treatment response after radiosurgery, focusing particularly on the heterogeneity of responses in a single patient, since it is often discussed in the context of immunotherapy of brain metastasis, but

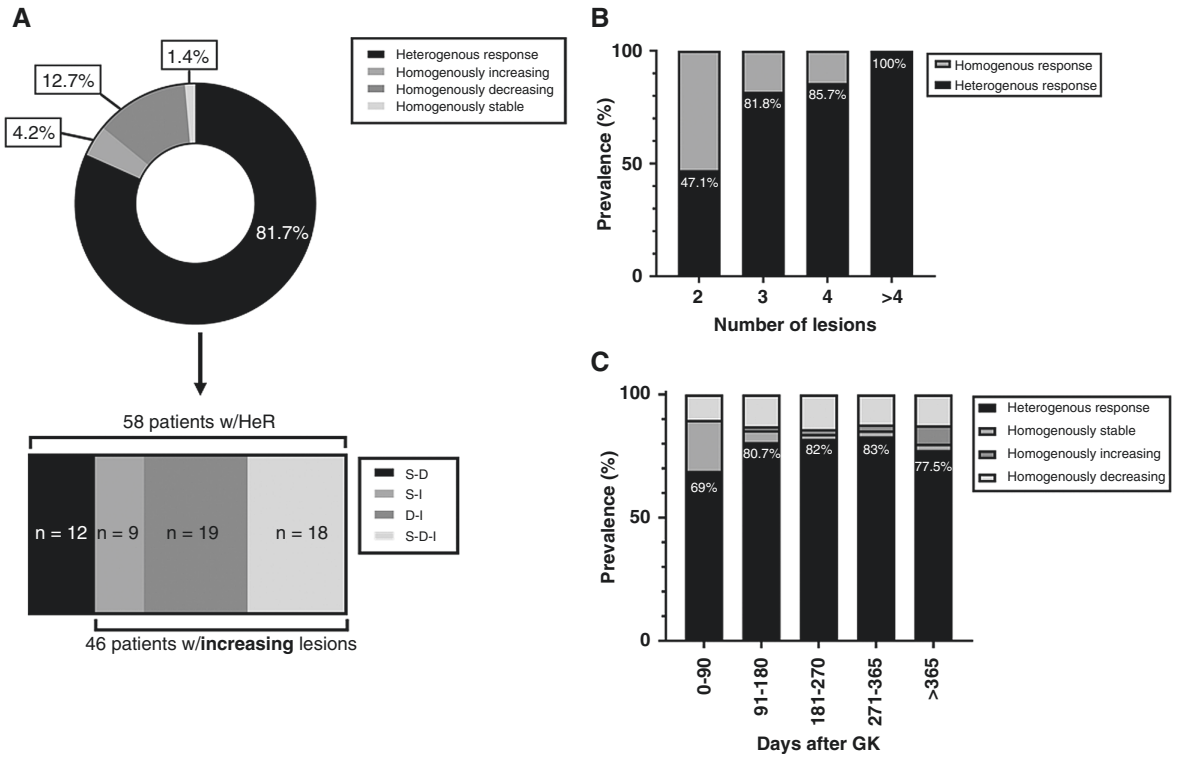


Figure 5. Distribution of lesions according to their response types: (A) Prevalence of different treatment responses at last date of follow-up among patients with >1 lesion. The lower graph dissects the group of heterogenous response (HeR), into the different components that make up HeR. S-D: stable and decreasing, S-I: stable and increasing, D-I: decreasing and increasing, S-D-I: stable, decreasing, and increasing lesions. (B) The chart indicates the prevalence of HeR in relation to homogenous response as a function of the total number of lesions per patient. (C) The graph shows the proportion of different treatment responses on follow-up scans as a function of the number of days after first GK intervention.

rarely in radiosurgery.[11–13] By considering patients with more than one lesion ($n = 71$), classifying the individual lesions as increasing, decreasing, or stable in size and finally comparing them we found that at the end of our follow-up 81.7% of patients showed a HeR, while HD, HI and HS were treatment response patterns found only in a minority of patients. Among those patients with a HeR, we also tried to understand which combinations of treatment responses were most prevalent. We found that the combinations of "decreasing and increasing," and "decreasing, stable, and increasing" lesions made up over half of the patients with HeR and were therefore the most common treatment responses in our patients at the end of follow-up. We also found that the prevalence of HeR was higher, the more metastatic lesions the patients had, reaching 100% in those who had more than four lesions. The importance of these findings lies in the way they underscore the need of implementing longitudinal individual lesion tracking and ensuring clear and detailed reporting of the changes in size of individual lesions by radiologists: while some metastatic lesions in a patient can increase in size over time, it is probable that not all will.[10] In other words, not all lesions will have to be treated, which is of especial importance for patients with a very large number of lesions. In this group, recognizing and tailoring the treatment to the smaller subset of lesions that are growing, might help spare patients from whole brain radiation treatment and other treatments such as surgical debulking of several lesions.[14] Up until now, treatment response is typically calculated by summation of diameters and comparing this summary measure to that in priors, yielding a decision on the progression, remission, or stability of the disease as a whole.[6,15] We suggest and encourage adding the description of "homogenous" or "heterogenous" response to the treatment response assessment to (i) acknowledge the different treatment needs of different lesions, and (ii) encourage radiologists to report disease change on a single lesion level, matching the high spatial resolution of nowadays treatment options. Furthermore, we encourage translational research into the mechanisms underlying HeR to SRS: As examined by other studies on other treatment modalities, HeR is likely a result of intratumoral geno- and phenotypic heterogeneity.[5,16] We believe that understanding precisely what genetic alterations render a metastasis more, or less resistant to SRS will help improve treatment protocols and clinical outcomes.

The newly developed LTT and our study have several limitations. First, the LTT works with either 1D or 2D lesion size measurements. While 2D lesion measurement is the current standard for lesion size measurements,[6,15] volumetric analysis has slowly begun to be implemented in research and clinical practice.[17–19] Accordingly, we aim to further develop this tool to incorporate 3D measurements and facilitate volumetric lesion growth tracking in the future. Furthermore, while the hanging protocol and lesion interface can reduce

the radiologist's workload substantially, the advent of automatic segmentation algorithms opens the possibility of incorporating their use into our workflow and enable fully automated volumetric tracking of brain lesions. Furthermore, our primary study included the assessment of lesions with a LD <5 mm at baseline, which is considered nonmeasurable by RANO-BM criteria. Nonetheless, acknowledging that punctate lesions are also treated with GK radiosurgery, we decided to include them but also perform a secondary analysis where these lesions were excluded. Importantly, both the primary and secondary studies showed a high prevalence of HeR at the end of follow-up and a trend of increasing proportions of HeR with a higher number of lesions. Finally, we utilized a number of clicks as an objective measure of efficiency and as a surrogate marker for the examiner's time and effort. However, as discussed above, the entire process of retrieving, reading, interpreting, and reporting the imaging results is greatly improved using LTT through streamlining of many additional steps, many of which are not easily measurable.

Stereotactic radiosurgical techniques such as GK have enabled the targeted treatment of brain metastases. In contrast, effective longitudinal assessment of single lesion size is often difficult in patients with multiple lesions and several prior images. Furthermore, effective communication of results is difficult with free-text reporting, which often leads to unclear and nonspecific reporting. We developed a PACS-integrated longitudinal LTT that circumvents these issues by offering an easy-to-use and efficient lesion measurement and tracking method, and automatically creates growth curves and tables that facilitate concise and objective result communication to treating clinicians. By longitudinally tracking over 490 lesions, we also show that lesions in a single patient respond to treatment heterogenously, particularly in cases with multiple lesions. These findings underscore the necessity for routine longitudinal tracking of lesions to ensure that treatment is tailored to the individual patient.

Supplementary material

Supplementary material is available at *Neuro-Oncology Advances* online.

Keywords

brain metastasis tracking | heterogenous response | lesion | radiosurgery | response to treatment.

Funding

American Society of Neuroradiology Fellow Award 2018 (MSA). This publication was made possible by KL2 TR001862 from the National Center for Advancing Translational Science, components of the National Institutes of Health (NIH), and NIH roadmap for Medical Research.

Conflict of interest statement. MingDe Lin is an employee and stockholder of Visage Imaging, Inc., and unrelated to this work, receives funding from NIH/NCI R01 CA206180. Khaled Bousabarah and Randy Owens are employee of Visage Imaging, Inc. Veronica Chiang is a paid consultant for Monteris Medical Inc.

Authorship statement. Conceptualization: G.C.P., K.B., B.J., S.M., R.O., A.O., V.C., M.L.; MSAData curation: G.C.P., L.J., M.S.A.; Formal analysis: G.C.P.; MSAFunding acquisition: M.S.A.; Investigation: G.C.P., K.B., M.v.R., B.J., S.A.F., I.I.; Methodology: G.C.P., V.C.; MSAProject administration: G.C.P., M.S.A.; Resources: A.O., V.C.; MSASoftware: K.B., R.O., M.L.;Supervision: M.S.A., M.L., V.C.; AOValidation: G.C.P., K.B., M.v.R., L.J., I.I.; Visualization: G.C.P., T.V., A.G., M.L.; Writing: original draft: G.C.P., K.B., T.V., A.G.; Writing: review & editing: M.S.A., M.v.R., L.J., S.M., M.L., A.O., V.C., S.A.F., I.I.

References

1. Nayak L, Lee EQ, Wen PY. Epidemiology of brain metastases. *Curr Oncol Rep.* 2012;14(1):48–54.
2. Chen JY, Vedantham S, Lexa FJ. Burnout and work-work imbalance in radiology-wicked problems on a global scale. A baseline pre-COVID-19 survey of US neuroradiologists compared to international radiologists and adjacent staff. *Eur J Radiol.* 2022:110153. doi:10.1016/j.ejrad.2022.110153.
3. Benson J, Burgstahler M, Zhang L, Rischall M. The value of structured radiology reports to categorize intracranial metastases following radiation therapy. *Neuroradiol J.* 2019;32(4):267–272.
4. Bink A, Benner J, Reinhardt J, et al. Structured reporting in neuroradiology: intracranial tumors. *Front Neurol.* 2018;9:32.
5. Crusz SM, Tang YZ, Sarker SJ, et al. Heterogeneous response and progression patterns reveal phenotypic heterogeneity of tyrosine kinase inhibitor response in metastatic renal cell carcinoma. *BMC Med.* 2016;14(1):185.
6. Lin NU, Lee EQ, Aoyama H, et al. Response assessment criteria for brain metastases: proposal from the RANO group. *Lancet Oncol.* 2015;16(6):e270–e278.
7. McTyre ER, Soike MH, Farris M, et al. Multi-institutional validation of brain metastasis velocity, a recently defined predictor of outcomes following stereotactic radiosurgery. *Radiother Oncol.* 2020;142(1):168–174.
8. Hayward JH, Ray KM, Wisner DJ, et al. Improving screening mammography outcomes through comparison with multiple prior mammograms. *Am J Roentgenol.* 2016;207(4):918–924.
9. Da Silva AN, Nagayama K, Schlesinger D, Sheehan JP. Early brain tumor metastasis reduction following Gamma Knife surgery. *J Neurosurg.* 2009;110(3):547–552.
10. Kerkhof M, Ganeff I, Wiggenraad RGJ, et al. Clinical applicability of and changes in perfusion MR imaging in brain metastases after stereotactic radiotherapy. *J Neurooncol.* 2018;138(1):133–139.
11. Ferrara R, Matos I. Atypical patterns of response and progression in the era of immunotherapy combinations. *Future Oncol.* 2020;16(23):1707–1713.
12. Jan Willem Rauwerdink D, van Persijn van Meerten E, van der Hage J, Kapiteijn E. Management of heterogeneous tumor response patterns to

immunotherapy in patients with metastatic melanoma. *Melanoma Res.* 2022;32(1):45–54.

13. Kwak JJ, Tirumani SH, Van den Abbeele AD, Koo PJ, Jacene HA. Cancer immunotherapy: imaging assessment of novel treatment response patterns and immune-related adverse events. *Radiographics.* 2015;35(2):424–437.
14. Shinde A, Akhavan D, Sedrak M, Glaser S, Amini A. Shifting paradigms: whole brain radiation therapy versus stereotactic radiosurgery for brain metastases. *CNS Oncol.* 2019;8(1):CNS27.
15. 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.
16. Humbert O, Chardin D. Dissociated response in metastatic cancer: an atypical pattern brought into the spotlight with immunotherapy. *Front Oncol.* 2020;10:566297.
17. Huang RY, Young RJ, Ellingson BM, et al. Volumetric analysis of IDH-mutant lower-grade glioma: a natural history study of tumor growth rates before and after treatment. *Neuro Oncol.* 2020;22(12):1822–1830.
18. Oft D, Schmidt MA, Weissmann T, et al. Volumetric regression in brain metastases after stereotactic radiotherapy: time course, predictors, and significance. *Front Oncol.* 2020;10:590980.
19. Patel RA, Lock D, Helenowski IB, et al. Postoperative stereotactic radiosurgery for patients with resected brain metastases: a volumetric analysis. *J Neurooncol.* 2018;140(2):395–401.