Neuro-Oncology

The RANO Leptomeningeal Metastasis Group proposal to assess response to treatment: lack of feasibility and clinical utility and a revised proposal

Emilie Le Rhun, Patrick Devos, Thomas Boulanger, Marion Smits, Dieta Brandsma, Roberta Rudà, Julia Furtner, Johann-Martin Hempel, Tjeerd J. Postma, Patrick Roth, Tom J. Snijders, Frank Winkler, Sebastian Winklhofer, Antonella Castellano, Elke Hattingen, Jaume Capellades, Thierry Gorlia, Martin van den Bent, Patrick Y. Wen, Martin Bendszus, and Michael Weller[○]; on behalf of the European Organisation for Research and Treatment of Cancer (EORTC) Brain Tumor Group (BTG) Central Nervous System (CNS) Metastases Committee and the EORTC BTG Imaging Committee

Department of Neurology & Brain Tumor Center, University Hospital and University of Zurich, Zurich, Switzerland (E.L.R., P.R., M.W.); University of Lille, France; Neuro-oncology, Neurosurgery Deparment, CHU Lille, France: Neurology, Medical oncology Department, Oscar Lambret Center, Lille, France (E.L.R.); University Lille, EA2694, France (P.D.); Imaging Department, Oscar Lambret Center, LIlle, France (T.B.); Department of Radiology and Nuclear Medicine, Erasmus MC–University Medical Center Rotterdam, Rotterdam, Netherlands (M.S.); Department of Neuro-oncology, Netherlands Cancer Institute–Antoni van Leeuwenhoek, Amsterdam, Netherlands (D.B.); Department of Neuro-oncology, City of Health and Science and University of Turin, Turin, Italy (R.R.); Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Vienna, Austria (J.F.); Radiology Department, Department for Diagnostic and Interventional Radiology, University of Tübingen, Tübingen, Germany (J.-M.H.); Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Neurology, Amsterdam, Netherlands (T.J.P); Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands (T.J.S.); Neurology Clinic and National Center for Tumor Diseases, University Hospital Heidelberg, Heidelberg, Germany, and Clinical Cooperation Unit Neuro-oncology, German Cancer Consortium, German Cancer Research Center, Heidelberg, Germany (F.W.); Department of Neuroradiology, University Hospital Zurich, University of Zurich, Zurich, Switzerland (S.W.); Neuroradiology Department, Vita-Salute San Raffaele University and IRCCS San Raffaele Scientific Institute, Milan, Italy (A.C.); Institute of Neuroradiology, University Hospital Frankfurt/Main, Frankfurt, Germany (E.H.); Neuroradiology Department, Hospital del Mar, Barcelona, Spain (J.C.); EORTC Headquarters, Brussels, Belgium (T.G.); Brain Tumor Center at Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, Netherlands (M.v.d.B.); Center for Neuro-Oncology, Dana-Farber/Brigham and Women's Cancer Center, Boston, Massachusetts, USA (P.Y.W.); Department of Neuroradiology, University of Heidelberg, Heidelberg, Germany (M.B.)

Corresponding Author: Prof Michael Weller, Department of Neurology and Brain Tumour Centre, University Hospital, Zurich, CH-8091 Zurich, Switzerland (michael.weller@usz.ch).

Abstract

Background. A scorecard to evaluate magnetic resonance imaging (MRI) findings during the course of leptomeningeal metastases (LM) has been proposed by the Response Assessment in Neuro-Oncology (RANO) group.

Methods. To explore the feasibility of the Leptomeningeal Assessment in Neuro-Oncology (LANO) scorecard, cerebrospinal MRIs of 22 patients with LM from solid tumors were scored by 10 neuro-oncologists and 9 neuroradiologists at baseline and at follow-up after treatment. Raters were blinded for clinical data including treatment. Agreement between raters of single items was evaluated using a Krippendorff alpha coefficient. Agreement between numerical parameters such as scores for changes between baseline and follow-up and total scores was evaluated by determining the intraclass coefficient of correlation.

Results. Most raters experienced problems with the instructions of the scorecard. No acceptable alpha concordance coefficient was obtained for the rating of single items at baseline or follow-up. The most concordant ratings were obtained for spinal nodules. The concordances were worst for brain linear leptomeningeal enhancement and cranial nerve enhancement. Discordance was less prominent among neuroradiologists than among

neuro-oncologists. High variability was also observed for evaluating changes between baseline and follow-up and for total scores.

Conclusions. Assessing response of LM by MRI remains challenging. Central imaging review is therefore indispensable for clinical trials. Based on the present results, we propose a new, simplified scorecard that will require validation using a similar approach as pursued here. The main challenges are to define measurable versus nonmeasurable (target) lesions and measures of change that allow assessment of response.

Key Points

1. Validated tools to diagnose and assess course of disease in leptomeningeal metastasis are lacking.

2. A panel of experienced neuroradiologists and neuro-oncologists failed to arrive at consensus ratings using the RANO LM score.

3. A novel leptomeningeal metastasis rating score based on the current survey is proposed.

Importance of the Study

A standardized approach is recommended for the assessment of response in LM; however, no such approach has been validated yet. We explore here the feasibility of the scorecard proposed by the RANO group. A panel of 10 neuro-oncologists and 9 neuroradiologists experienced challenges not only with the instructions of the scorecard, but also in arriving at concordant ratings for many items. We conclude that the present scorecard is not adequate for clinical practice or for clinical trials, but needs to be simplified, improved, and again validated.

Five to ten percent of solid cancer patients will present with leptomeningeal metastases (LM) during the course of their disease. Survival is limited to only 10% of patients alive at one year across solid cancer entities.[1–5] Response Assessment in Neuro-Oncology (RANO) formed a Leptomeningeal Assessment in Neuro-Oncology (LANO) working group that generated a proposal on the response assessment in LM.[6] The first European Association of Neuro-Oncology (EANO) and European Society for Medical Oncology (ESMO) guidelines on LM[7] aimed at further standardizing diagnosis and therapeutic approaches. The diagnosis of LM and response assessment during follow-up depend on clinical examination, cerebrospinal imaging, and standard cerebrospinal fluid (CSF) cytology. Typical symptoms and signs of LM include headache; nausea and vomiting; mental changes; gait difficulties; cranial nerve palsies, such as with diplopia or visual disturbance (cranial nerves II, III, IV, VI) and hearing loss (cranial nerve VIII); radicular signs including weakness, voiding, and cauda equina problems; and focal or radicular neck and back pain. Characteristic MRI findings include leptomeningeal enhancement or linear ependymal enhancement, cranial and spinal nerve root enhancement, and leptomeningeal enhancing nodules, notably of the cauda equina. CSF cytological analysis should be reported as (i) positive, defined as the presence of malignant cells in the CSF; (ii) equivocal, corresponding to the detection of suspicious or atypical cells in the CSF; or (iii) negative, defined as the absence of malignant or potentially malignant (suspicious or equivocal) cells in the CSF.

A standardized scorecard to aid in the evaluation of MRI findings during the course of disease has been proposed by the LANO group. In this scorecard, LM main features and the different types of CNS metastases shall be reported as present or absent, dimensions of measurable nodules can be noted, and changes from the previous MRI shall be scored from –3 to +3 (Supplementary Table 1, left panel).[6] We explore here the feasibility of this scorecard for the evaluation of response in 22 patients treated for LM.

Materials and Methods

MRI Evaluation

Anonymized paired MRI scans from 22 adult patients with LM from breast cancer, lung cancer, or melanoma treated between baseline and follow-up evaluation at Oscar Lambret Cancer Center, Lille, France, were used. All except 1 patient were dead at the time of the project; the surviving patient provided written consent. The scans were provided to each participant with a copy of the LANO scorecard and its instructions as published in the original manuscript,[6] without other recommendations to avoid bias in the interpretation. Brain MRI included unenhanced axial T1-weighted, axial fluid attenuated inversion recovery, and axial diffusion-weighted and gadolinium-enhanced T1-weighted 3D sequences. Spinal MRI included sagittal T2-weighted and unenhanced as well as contrast-enhanced T1-weighted sequences, supplemented by axial images as needed. Participating raters were senior neuro-oncologists (n = 10) or senior neuroradiologists (n = 9), who are all authors of this article. They were blinded to clinical characteristics and treatments received

by the patients. The ratings were centrally collected and analyzed by E.L.R. and M.W.; the primary statistical analysis was done by P.D. This study was conducted on behalf of the EORTC BTG CNS Metastases Committee and the EORTC BTG Imaging Committee.

Statistical Analysis

Binary variables are described with frequencies, continuous parameters with medians and quartiles. Interrater agreement for binary variables (such as absence or presence of a feature on the MRI) was assessed using the Krippendorff α statistic.[8] Values close to 1 indicate high interrater agreement, whereas values close to 0 indicate that agreement in raters' measurements of that feature for the same patient is by chance. An alpha coefficient of ≥ 0.8 is usually considered reliable, an alpha coefficient of $0.800 >$ alpha ≥ 0.667 may be acceptable, and measures with alpha <0.667 show disagreement. Interrater agreement between numerical parameters such as scores for changes between baseline and follow-up and of the total scores was evaluated by determining the intraclass correlation coefficient (ICC).[9] An ICC >0.8 indicates good agreement. In order to detect discordant raters who could lower the agreement, we computed measures of influence. There is no gold standard to address this question and we do not know the "true" value of a parameter. For binary parameters, the reference value for a patient was determined as 1 if at least 10 raters among the 19 rated 1; 0 otherwise. So, for each rater, we can compute a score from 0 to 440 (20 parameters and 22 patients) corresponding to the number of items coded the same as the reference value. For numerical parameters, reference values for a patient were computed as the median of the values of the 19 raters. For each rater, we were thus able to compute a distance from the reference values. Box-plot representation was used to detect abnormal raters. Analyses were performed with SAS v9.4 and SPSS 22.

This study was approved under the number 2018-00192 by the Cantonal Ethics Committee Zurich.

Results

Characteristics of Patients and MRI

LM was confirmed in 18 cases, probable in 2 cases, and possible in 2 cases according to EANO ESMO guidelines. Primary tumors were breast cancer ($n = 16$), melanoma ($n = 3$), and lung cancer ($n = 3$). Median age at LM diagnosis was 58.5 years (range 31–74). Ten patients had been treated for CNS metastases prior to the diagnosis of LM. Therapeutic measures between the 2 MRI assessments included systemic treatment alone in 7 patients, a combination of systemic and intra-CSF pharmacotherapy in 11 patients, a combination of systemic treatment and focal radiotherapy in 1 patient, and a combination of systemic and intra-CSF treatment and focal radiotherapy in 3 patients. No whole brain radiotherapy was administered. Median overall survival from LM diagnosis was 193 days (range 56–1121) (Supplementary Table 2). The median interval between brain and spinal MRI at baseline was 1 day (range 0–23); the 2 examinations were performed on the same day in 11 patients. The median interval between the first MRI at baseline and the first MRI during follow-up

was 2.7 months (range 0.82–9.1 mo). Follow-up brain and spinal MRI were performed on the same day for 20 patients (Supplementary Table 3).

Feasibility of the LANO Scorecard

After receipt of the completed scorecards and comments, we observed that the instructions for columns 2 and 3, where the different LM characteristics should be scored as present/absent/non-evaluable and measured, were interpreted differently between raters. Especially, it was not clear whether these columns referred to the baseline or the follow-up MRI evaluation. All raters were then asked to complete 2 more columns, with evaluation of the LM characteristics at baseline and at first evaluation after treatment, for clarification (Supplementary Table 1, right panel).

Rating of Single Items of the LANO Scorecard

Supplementary Table 4 summarizes the assessment of the 10 items of the baseline MRI by the raters. A high variability was observed when evaluating the median values, ranges, and quartiles of the absence or presence of single items (Figure 1A, Supplementary Table 5). For instance, for leptomeningeal nodules in the brain, the most conservative rater saw these in only 6 patients (27%), whereas the least conservative rater saw them in 19 patients (86%). There was no acceptable interobserver agreement, since all the values of the alpha Krippendorff coefficient were inferior to 0.67, both for all raters pooled and for neuro-oncologists and neuroradiologists considered separately (Table 1). Among LM-relevant items, the best interobserver concordance was seen for the evaluation of spinal leptomeningeal nodules and for spinal linear leptomeningeal and nerve root enhancement, with alpha coefficients of 0.60, 0.45, and 0.45. Yet, spinal nodules were noted by only 25% of the raters, thus, in most cases, the raters agreed on the absence of spinal nodules. The majority of raters diagnosed spinal nodules in patient (P) 13 and P22 (Figure 1B, Supplementary Table 4). Poor interobserver agreement was obtained for brain-related items, with alpha coefficients of 0.40–0.41. Among non-LM-related items, the best agreement was observed for intramedullary metastases, with an alpha coefficient of 0.57. Yet, intramedullary metastases were scored as absent by 91% of the raters. The alpha coefficients for epidural metastases, brain metastases, and hydrocephalus were below 0.50. Expectedly, similar results were obtained when the rating of single items of the follow-up examination were analyzed (data not shown).

The raters were also asked to determine the size of any measurable LM nodules using maximum orthogonal diameters. Supplementary Table 6 shows that most nodules were considered nonmeasurable, with the exception of P6, P18, and P21 for brain nodules and of P3, P11, and P22 for spinal nodules, which were rated measurable by more than 50% of the raters.

Deriving Total Sum Scores from the LANO Scorecard

Next we analyzed the total sum scores of the LANO scorecard. To explore the validity and the clinical usefulness of
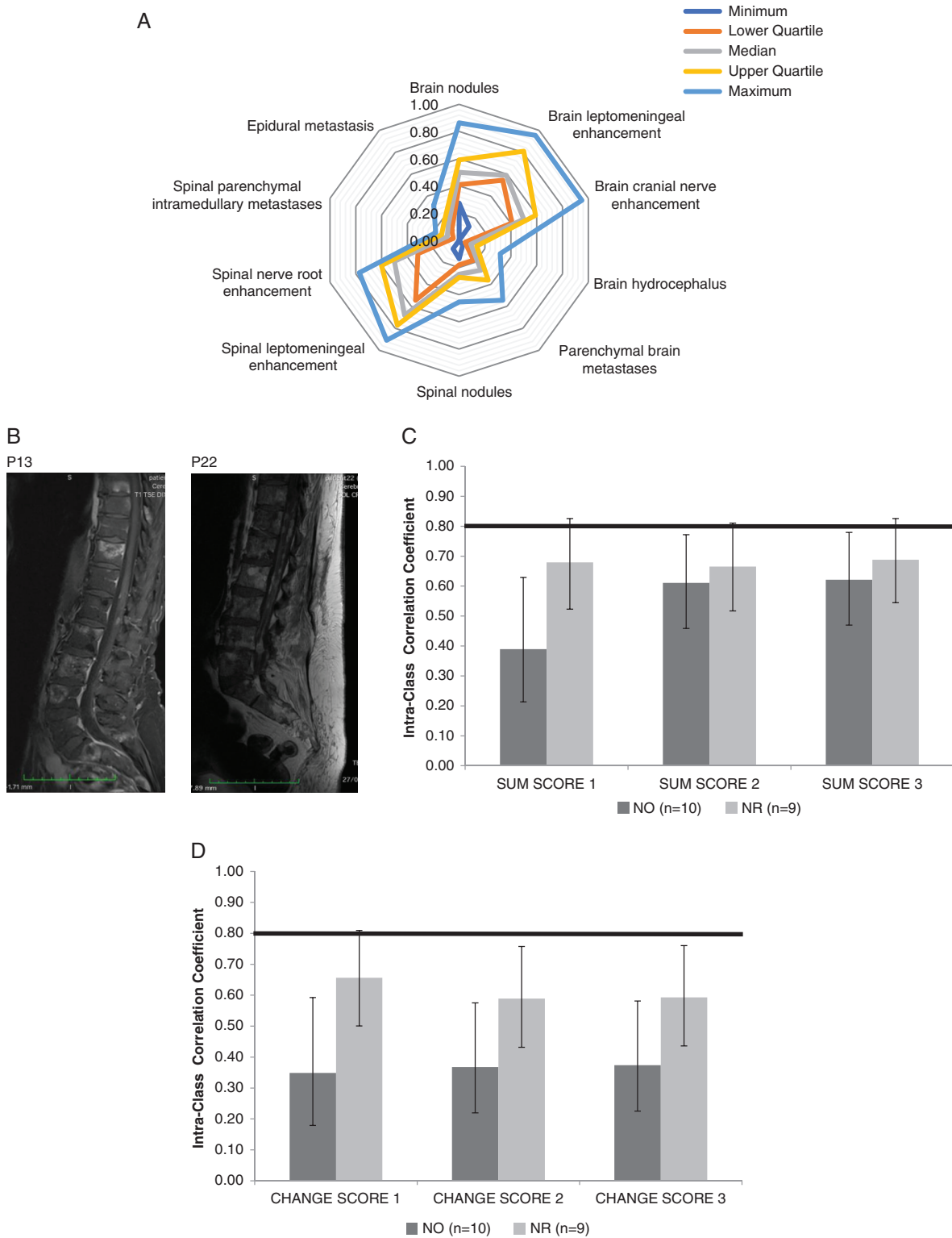
Fig. 1 Neuroradiological assessment of LM using the LANO scorecard. (A) Median, range, and quartiles for interrater variability of single RANO items upon assessment of newly diagnosed LM at baseline assessment. (B) Representative sagittal spinal contrast-enhanced, T1-weighted MRI from P13 and P22, rated as showing nodules by a majority of raters. (C, D) Overall concordance rates for neuro-oncologists (NO) versus neuroradiologists (NR) for the sum scores (C) and the change scores (D) (for details see Tables 2 and 4).

Table 1 Overall concordance for single items of the LANO grid at baseline (Krippendorff alpha coefficient)*

MRI Findings	All Raters (NO + NR) (n = 19)			NO (n = 10)			NR (n = 9)		
	Alpha	Lower Limit 95% CI	Upper Limit 95% CI	Alpha	Lower Limit 95% CI	Upper Limit 95% CI	Alpha	Lower Limit 95% CI	Upper Limit 95% CI
Brain									
Nodules (subarachnoid or ventricular)	**0.40**	0.22	0.58	**0.47**	0.30	0.64	**0.34**	0.16	0.50
Leptomeningeal enhancement	**0.41**	0.22	0.59	**0.50**	0.32	0.68	**0.35**	0.14	0.54
Cranial nerve enhancement	**0.40**	0.22	0.58	**0.39**	0.21	0.57	**0.47**	0.29	0.63
Hydrocephalus	0.43	0.05	0.70	0.31	-0.09	0.66	0.59	0.30	0.83
Parenchymal (brain metastases)	0.48	0.28	0.67	0.53	0.34	0.72	0.43	0.23	0.62
Spine									
Nodules (subarachnoid)	**0.60**	0.39	0.78	**0.55**	0.32	0.75	**0.62**	0.44	0.81
Leptomeningeal enhancement	**0.45**	0.26	0.63	**0.43**	0.22	0.61	**0.45**	0.27	0.60
Nerve root enhancement	**0.45**	0.26	0.64	**0.51**	0.34	0.68	**0.39**	0.20	0.56
Parenchymal (intramedullary metastases)	0.57	0.23	0.87	0.60	0.26	0.92	0.54	0.22	0.79
Epidural metastasis	0.46	0.16	0.69	0.43	0.11	0.72	0.52	0.25	0.75

Abbreviations: NO: neuro-oncologists, NR: neuroradiologists.

*An alpha coefficient of ≥0.8 is usually considered reliable, an alpha coefficient of 0.800 > alpha >0.667 may be acceptable, and measures with alpha <0.667 show disagreement. LM-related items are printed in bold.

Table 2 Overall concordance of the total score at baseline using ICC

Item	All Raters (NO + NR) (n = 19)			NO (n = 10)			NR (n = 9)		
	ICC	LCL	UCL	ICC	LCL	UCL	ICC	LCL	UCL
Sum score 1	0.43	0.26	0.65	0.39	0.21	0.63	0.68	0.52	0.83
Sum score 2	0.63	0.49	0.78	0.61	0.46	0.77	0.67	0.52	0.81
Sum score 3	0.65	0.51	0.79	0.62	0.47	0.78	0.69	0.54	0.83

Abbreviations: LCL: lower 95% confidence limit, UCL: upper 95% confidence limit.
Sum score 1: as rated by the participants without editing.
Sum score 2: as corrected according to LANO proposal with 7 items (brain nodules, brain leptomeningeal enhancement, cranial nerve enhancement, spinal nodules, spinal leptomeningeal enhancement, nerve root enhancement, and epidural metastases).
Sum score 3: as corrected with 6 leptomeningeal items only (brain nodules, brain leptomeningeal enhancement, cranial nerve enhancement, spinal nodules, spinal leptomeningeal enhancement, and nerve root enhancement), excluding epidural metastases.

the sum scores, we had to introduce modifications into the analysis. First, we report the "real life" sum scores as provided by the raters and excluded raters who had not provided a sum score; 6 raters did not provide a sum score for at least one patient (sum score 1). Second, we recalculated the scores for all raters including only the 7 items as proposed by the LANO group (brain nodules, brain leptomeningeal enhancement, cranial nerve enhancement, spinal nodules, spinal leptomeningeal enhancement, nerve root enhancement, and epidural metastases) (sum score 2). Third, we calculated a 6-item score omitting epidural lesions because these should not be part of LM assessment (sum score 3) (Supplementary Table 7). The agreement between raters for the sum scores is shown in Table 2 and Figure 1C. The ICC was always below 0.8, indicating poor agreement.

Rating of Change for Single Items of the LANO Scorecard

We then looked at the change between baseline and first evaluation for each item of the LANO scorecard. Supplementary Table 8 indicates how the change from baseline was rated for each item of the LANO scorecard. The agreement between the raters was again low (Table 3). When considering the LM items, the best results were obtained for cranial nerve enhancement (ICC = 0.45), which still reflects very poor agreement. When considering non-LM items, the highest agreement was obtained for brain metastases (ICC = 0.54), whereas agreements for hydrocephalus, epidural metastases, and especially intramedullary changes were low, with a negative value of ICC suggesting actual disagreement among raters. A comparison of neuro-oncologists and neuroradiologists indicated less agreement among neuro-oncologists, but also the neuroradiologists never achieved an ICC of 0.8.

Deriving Composite Scores of Change from the LANO Scorecard

Composite scores of change (total change score 1) were always filled in by 15 raters, never filled in by 3 raters, and 1 rater omitted one case. To explore the validity and the clinical usefulness of the composite scores, we applied the same modifications as in Supplementary Table 7 to derive scores of change for LM-related items based on 7 respectively 6 items. The mean values and ranges of the 3 composite change scores are reported in Supplementary Table 9 and show again strong variability among raters. This was confirmed when evaluating the interobserver agreement by calculating the ICC, which never reached 0.8 for any of the composite change scores, but confirmed the trend for more agreement among neuroradiologists (Table 4A, Figure 1D). We also explored whether specific patients or raters were largely responsible for the poor interrater agreement. This was not the case for specific patients (data not shown), but omission of ratings from 2 "outlier" neuro-oncologists resulted in much better agreement, but still not to an acceptable level (Table 4B).

Assessing Response Using the LANO Scorecard

According to the LANO publication,[6] progressive disease is defined by a 25% worsening of the score between baseline and follow-up, whereas a partial response corresponds to a 50% improvement of the score, a resolution of all baseline MRI abnormalities to a complete response, and all other situations to stable disease. However, the total LANO score in column 4 (Supplementary Table 1, left), which can assume values between −21 and +21, represents a score that summarizes relative changes. A relative change score cannot be put into a percentage relation to an absolute score at baseline, and the LANO recommendations provide no guidance of how to derive such a score. Accordingly, progressive disease or partial response cannot be determined using this approach.

Discussion

Among various initiatives of the RANO working group, a recommendation for MRI-based diagnostic assessment and response evaluation in LM based on a dedicated scorecard has been developed.[6] A literature survey indicated that this score has not been used in publications on LM and we assume that it has not been introduced into clinical practice either. Accordingly, here we sought to explore the feasibility of this scorecard by requesting 10 neuro-oncologists and 9 neuroradiologists to rate 22 paired MRI series of patients

Table 3 Overall concordance rates for the changes of RANO single items upon assessment of newly diagnosed LM between first evaluation and baseline evaluated using ICC

MRI Findings	All Raters (NO + NR) (n = 19)			NO (n = 10)			NR (n = 9)		
	Alpha	Lower Limit 95% CI	Upper Limit 95% CI	Alpha	Lower Limit 95% CI	Upper Limit 95% CI	Alpha	Lower Limit 95% CI	Upper Limit 95% CI
Brain									
Nodules (subarachnoid or ventricular)	**0.37**	0.23	0.58	**0.27**	0.13	0.50	**0.42**	0.26	0.62
Leptomeningeal enhancement	**0.34**	0.20	0.57	**0.30**	0.16	0.53	**0.32**	0.17	0.54
Cranial nerve enhancement	**0.45**	0.28	0.69	**0.19**	0.07	0.42	**0.68**	0.52	0.83
Hydrocephalus	0.09	0.03	0.22	0.11	0.02	0.28	0.06	-0.03	0.22
Parenchymal (brain metastases)	0.54	0.37	0.75	0.45	0.28	0.66	0.57	0.39	0.76
Spine									
Nodules (subarachnoid)	**0.35**	0.21	0.56	**0.23**	0.10	0.45	**0.52**	0.35	0.72
Leptomeningeal enhancement	**0.17**	0.06	0.42	**0.34**	0.18	0.58	**0.43**	0.24	0.70
Nerve root enhancement	**0.31**	0.17	0.57	**0.17**	0.05	0.40	**0.44**	0.27	0.67
Parenchymal (intramedullary metastases)	-0.04	-0.05	-0.02	-0.10	-0.11	-0.09	0.08	-0.02	0.29
Epidural metastasis	0.04	-0.01	0.15	0.05	-0.02	0.21	0.28	0.12	0.52

Abbreviations: NO: neuro-oncologist. NR: neuroradiologist. LM-related items are printed in bold.

Table 4 Overall concordance rates for the 3 composite change scores determined using ICC

A. Including All Raters									
	NO + NR (*n* = 19)			**NO (*n* = 10)**			**NR (*n* = 9)**		
Item	**ICC**	**LCL**	**UCL**	**ICC**	**LCL**	**UCL**	**ICC**	**LCL**	**UCL**
Change score 1	0.48	0.32	0.70	0.35	0.18	0.59	0.66	0.50	0.81
Change score 2	0.48	0.33	0.66	0.37	0.22	0.58	0.59	0.43	0.76
Change score 3	0.48	0.34	0.67	0.37	0.23	0.58	0.59	0.44	0.76
B. Excluding the 2 Outlier Raters									
	NO + NR (*n* = 17)			**NO (*n* = 8)**					
Item	**ICC**	**LCL**	**UCL**	**ICC**	**LCL**	**UCL**			
Change score 1	0.60	0.44	0.79	0.58	0.39	0.78			
Change score 2	0.57	0.42	0.74	0.55	0.39	0.74			
Change score 3	0.58	0.43	0.75	0.56	0.40	0.74			

Abbreviations: ICC intraclass correlation, LCL, lower 95% confidence limit, UCL: upper 95% confidence limit.
Total score 1: as rated by the participants.
Total score 2: as corrected according to LANO proposal with 7 items (brain nodules, brain leptomeningeal enhancement, cranial nerve enhancement, spinal nodules, spinal leptomeningeal enhancement, nerve root enhancement, and epidural metastases).
Total score 3: as corrected with the 6 leptomeningeal items only (brain nodules, brain leptomeningeal enhancement, cranial nerve enhancement, spinal nodules, spinal leptomeningeal enhancement, and nerve root enhancement).

with diagnosed LM (baseline) and subsequently treated for LM (follow-up scan). The study was set up in a naturalistic way so as to simulate introduction of this scoring system into a clinical trial or clinical routine.

Neither the neuro-oncologists nor the neuroradiologists achieved adequate interobserver agreement. Among the LM-related items, spinal, although not brain, nodules were scored more consistently than the items related to linear contrast enhancement, although participants mostly agreed on the absence rather than presence of nodules (Table 1). Since the LANO scorecard asks for the absence or presence of nodules without defining a nodule, the raters assessed for nodules without guidance of their definition. We noted, however, that the vast majority of nodules were considered nonmeasurable and were thus likely smaller than 5 × 10 mm (Supplementary Table 6). Even when the total scores of change were edited upon central review of all feedback to correct for erroneous consideration of non-LM-related items according to the original LANO score and when epidural metastases were excluded, corresponding to sum score 3, there was still poor agreement (Figure 1C, D). We noted a trend to more consistent ratings provided by neuroradiologists compared with neuro-oncologists, but even neuroradiologists remained far from the threshold level of agreement of 0.8. Still, this trend may be interpreted to signify the need for professional (neuroradiologist) assessment of neuroimaging in LM.

The results of this effort show that the current scorecard is not useful, neither for current clinical practice nor for clinical trial conduct. Some instructions are apparently not sufficiently clear, or the scorecard is too complicated, or both (Supplementary Table 1). Specific challenges were (i) to understand that the form should be used to rate the current MRI and to compare it with the previous one, (ii) to use the proposed rating with "minus" or "plus" options to assess the change, and (iii) to derive a sum score that does not take into consideration (per instruction) changes for the items "hydrocephalus," "brain metastases," and "parenchymal medullary metastases."

In addition to the apparent challenges for experienced raters (the authors of this manuscript) to use the LANO scorecard instructions without further instructions, we identified additional weaknesses that should be overcome in a revised version of a structured tool for LM. These include elimination of epidural metastases from response assessment, the definition of a nodule, the distinction of leptomeningeal versus parenchymal brain disease, and parallel but clearly separate criteria to document brain parenchymal disease. Table 5 summarizes a proposal for a revised, greatly simplified scorecard that will require an iteration of the validation process conducted here for the prior version.

This study has limitations by its design. The raters were not specifically trained to use the LANO scorecard, but we preferred this approach to simulate clinical practice and base assessment only on the original instructions (rather than on interpretations provided by some of us). Inescapably, we conclude that the imaging response assessment of LM remains challenging and that the first LANO effort is not suitable to allow disease assessment in clinical practice or in clinical trials. We have incorporated the pitfalls and weaknesses observed with the current scorecard into the design of a new scorecard that will now undergo a similar process of evaluation. Meanwhile, we conclude that central imaging review in real time is essential for clinical trials in LM that involve imaging-based endpoints. The involvement of dedicated neuroradiologists at each clinical trial site is highly recommended and personal training to use MRI evaluation guidelines represents an option until more suitable standardized criteria have been developed.

Table 5 Proposal for a revised LANO grid

	Reference Scan	Follow-Up	Response Assessment[8]
Patient Identification Name or number: Sex: Date of birth:			
Dates of MRI			
Relevant history	n.a.	Treatment since reference scan[7]: Date of last lumbar puncture:	n.a.
MRI findings	Present (1) or absent (0) or non-evaluable (NE)[5] — Individual dimensions (N1, N2, N3: X x Y mm) of 3 largest measurable nodules (measurable defined as > 5 x 5 mm (orthogonal diameters in 2 planes)	Present (1) or absent (0) or non-evaluable (NE) — Individual dimensions (N1, N2, N3: X x Y mm) of 3 largest measurable nodules (measurable defined as > 5 x 5 mm (orthogonal diameters in 2 planes)	Change from previous MRI
Items related to assessment to leptomeningeal metastasis			
Brain			
Nodules (subarachnoid or ventricular)[1]			□ improved □ no change □ worse
Leptomeningeal linear enhancement[2]	n.a.[6]	n.a.	□ improved □ no change □ worse
Hydrocephalus[3]	n.a.	n.a.	□ improved □ no change □ worse
Spine			
Nodules (subarachnoid)			□ improved □ no change □ worse
Leptomeningeal linear enhancement	n.a.	n.a.	□ improved □ no change □ worse
Overall response assessment for LM (CR, PR, SD, PD, or NE)			
Items *not* related to assessment of leptomeningeal metastasis[4]			
Brain			
Parenchymal (brain) metastases			□ CR[9] □ PR □ SD □ PD

Table 5   Continued

Spine

Parenchymal (intramedul-
lary) metastases

☐ CR
☐ PR
☐ SD
☐ PD

Abbreviations: CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease.
Explanations:

[1]A **nodule** is a contrast-enhancing lesion that is defined as LM-related as opposed to parenchymal if there is direct contact (less than 2 mm distance) between the outer edge of the nodule and the leptomeninges on contrast-enhanced scans.

[2]**Leptomeningeal linear enhancement** may include cranial nerve or spinal nerve root, cerebellar folia, ventricular ependymal, or cerebral sulcal enhancement.

[3]**Hydrocephalus** is assessed by determining the Evans index calculated on T1-weighted axial MR images. It represents the ratio of the largest diameter at the maximal width of the frontal horns relative to the largest internal diameter of the cranium on the same slide (Brix et al. Eur J Radiol 2017;95:28–32).

[4]These items shall be documented as present or absent, but are not used for LM response assessment.

[5]NE refers to scans that cannot be assessed for poor quality or incomplete sequences.

[6]Not applicable.

[7]**Therapeutic options** for LM that should be noted here include any neurosurgical intervention, radiotherapy with information on the target of irradiation, systemic pharmacotherapy, and intrathecal pharmacotherapy. Assessing response treatment requires precise information on treatment delivered.

[8]**Progression** is diagnosed if there is at least one new measurable nodule, if at least one measurable nodule that does not reach 10 mm in its two largest perpendicular diameters, increases in the product of the largest perpendicular diameters by 50% or more, if at least one nodule of at least 10 mm diameter in its perpendicular diameters increases in the product of the largest perpendicular diameters by 25% or more, or if the largest ventricular diameter increases by at least 25%. *De novo* linear leptomeningeal contrast enhancement alone also qualifies for progression unless attributable to lumbar puncture. **Partial response** requires regression of the size of all measurable nodules by 50% or more, without an increase in ventricular size. **Complete response** requires resolution of all contrast-enhancing, LM-related measurable lesions, without an increase in ventricular size. All other situations are considered **stable disease.** LM without measurable nodules can only remain stable as its best response. Linear enhancement cannot be quantified and is thus only noted as absent or present, but not used for response assessment unless developing *de novo* or affecting the leptomeninges in anatomic regions not previously affected—then this constitutes **progressive disease.** Deterioration in any one item qualifying for progression will be sufficient to call progression.

[9]According to RANO imaging criteria for brain metastasis (Lin et al. Lancet Oncology 2015;16:e270–8).

Technical considerations

MRI scans should be performed on the same scanner or at least a device of identical field strength during follow-up using the same imaging protocol at all timepoints during the follow-up, gadolinium-based contrast agent should be injected ideally 10 min, but not less than 5 min before acquisition of T1-weighted sequences and the slice thickness should be 1 mm or less in the brain and 3 mm or less for the spinal cord, as the leptomeningeal enhancement may have complex aspects and is commonly linear (Le Rhun et al. Ann Oncol 2017;28:iv84–iv99).

Since lumbar punctures may induce leptomeningeal enhancement, the date(s) of the last CSF analysis performed before MRI acquisition should appear on the grid.

Supplementary Material

Keywords

Funding

References

1. Hyun J-W, Jeong IH, Joung A, Cho HJ, Kim S-H, Kim HJ. Leptomeningeal metastasis: clinical experience of 519 cases. *Eur J Cancer Oxf Engl 1990.* 2016;56:107–114.

2. Brower JV, Saha S, Rosenberg SA, Hullett CR, Ian Robins H. Management of leptomeningeal metastases: prognostic factors and associated outcomes. *J Clin Neurosci.* 2016;27:130–137.

3. Abouharb S, Ensor J, Loghin ME, et al. Leptomeningeal disease and breast cancer: the importance of tumor subtype. *Breast Cancer Res Treat.* 2014;146(3):477–486.

4. Kuiper JL, Hendriks LE, van der Wekken AJ, et al. Treatment and survival of patients with EGFR-mutated non-small cell lung cancer and leptomeningeal metastasis: a retrospective cohort analysis. *Lung Cancer.* 2015;89(3):255–261.

5. Geukes Foppen MH, Brandsma D, Blank CU, van Thienen JV, Haanen JB, Boogerd W. Targeted treatment and immunotherapy in leptomeningeal metastases from melanoma. *Ann Oncol.* 2016;27(6):1138–1142.

6. Chamberlain M, Junck L, Brandsma D, et al. Leptomeningeal metastases: a RANO proposal for response criteria. *Neuro Oncol.* 2016. doi: 10.1093/neuonc/now183.

7. 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 Off J Eur Soc Med Oncol.* 2017;28(suppl_4):iv84–iv99.

8. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures.* 2007;1:77–89.

9. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86(2):420–428.