This information is current as of October 11, 2024.

Comparison of Measurement Techniques and Response Criteria for MR Imaging Follow-up in Adult Primary Central Nervous System Lymphoma

K. Massicotte-Tisluck, D. Vanderweyen, J.-F. Vendrell, D. Fortin and G. Gahide

AJNR Am J Neuroradiol 2021, 42 (8) 1415-1420
doi: https://doi.org/10.3174/ajnr.A7132
http://www.ajnr.org/content/42/8/1415

Comparison of Measurement Techniques and Response Criteria for MR Imaging Follow-up in Adult Primary Central Nervous System Lymphoma

K. Massicotte-Tisluck, D. Vanderweyen, J.-F. Vendrell, D. Fortin, and G. Gahide

ABSTRACT

BACKGROUND AND PURPOSE: Current guidelines proposed for the measurement of primary central nervous system lymphoma in 2005 have indicated that unidimensional and bidimensional measurements may be used, using the same threshold for response categorization, because no clinical study has evaluated the agreement among the measurement techniques. Hence, our study assessed the agreement among different measurements.

MATERIALS AND METHODS: In this retrospective study, primary central nervous system lymphoma lesions were measured with different techniques (longest 1D, axial 1D, 2D, 3D, and the Response Evaluation Criteria in Solid Tumor) on consecutive MR images. Intra- and interobserver correlations were calculated with intraclass correlation coefficients. Correlations between raw measurements and variations in size compared with baseline were evaluated with the Spearman rank correlation, and agreement among response categories was evaluated.

RESULTS: A total of 304 examinations obtained in 40 patients was assessed. The intraobserver intraclass correlation coefficient for 3D, 2D, and longest 1D were ≥0.993. The interobserver intraclass correlation coefficient was ≥0.967. The correlations in raw measurements and size variation in comparison with 3D were respectively; 0.99 and 0.98 for 2D; 0.94 and 0.92 for longest 1D; 0.94 and 0.83 for axial 1D; and 0.90 and 0.79 for Response Evaluation Criteria in Solid Tumor. With 20%–30% and 25%–50% thresholds for unidimensional techniques, response categorizations were 95% and 95% for 2D, 92.5% and 90% for the longest 1D, 87.5% and 82.5% for axial 1D, and 90% and 85% for the Response Evaluation Criteria in Solid Tumor.

CONCLUSIONS: Both longest 1D and 2D demonstrated excellent correlations with 3D measurements. The longest 1D could be used for the follow-up of primary central nervous system lymphoma. If unidimensional measurements were used, 20% and 30% cutoffs should be used for defining response categorization instead of the current guidelines.

ABBREVIATIONS: CNS = central nervous system; CR = complete response; ICC = intraclass correlation coefficient; PCNSL = primary central nervous system lymphoma; PD = progressive disease; PR = partial response; RECIST = Response Evaluation Criteria in Solid Tumor; SD = stable disease; WBRT = whole brain radiotherapy

Since the first international guidelines were published in 1981 by the World Health Organization,[1] there are currently 2 main coexisting paradigms for clinically assessing tumor responses to treatments. On the one hand, there are unidimensional criteria, mainly represented by the Response Evaluation Criteria in Solid Tumor (RECIST). These criteria are widely accepted for the assessment of nonneurologic tumors. On the other hand, bidimensional criteria are mainly used for assessing primary cerebral tumors, such as the modified Macdonald criteria and the Response Assessment in Neuro-Oncology Criteria for primary CNS tumors.[2-4] The unidimensional criteria are characterized by partial response (PR), defined as a 30% decrease in size, and progressive disease (PD), defined as a 20% increase in size. The bidimensional criteria are slightly different, with PR defined as a 50% decrease in size and PD corresponding to a 25% increase. The currently accepted guidelines for assessing primary central nervous system lymphomas (PCNSLs) indicate that unidimensional or bidimensional criteria may both be used indiscriminately.[5] Most interesting, the recommended response criteria proposed by the authors were identical, notwithstanding the chosen measurement technique, ie, a 50% decrease in size to define PR and a 25% increase in size to define PD for both unidimensional and bidimensional measurements.

Received November 22, 2020; accepted after revision February 19, 2021.

From the Département d'Imagerie Médicale, (K.M.-T., D.V., G.G.) and Service de Neurochirurgie, Département de Chirurgie (D.F.), Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Québec, Canada; and Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke, Etienne Le Bel (D.F., G.G.), Faculté de Médecine et des Sciences de la Santé (K.M.-T., D.V.), Université de Sherbrooke, Sherbrooke, Québec, Canada and Departments of Neuroradiology (J.-F.V.), Institut de cancérologie de Montpellier, Clinique de val d Aurelle, Montpellier, France.

Please address correspondence to Gérald Gahide, MD, Radiology Department, Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Quebec, Canada. 3001 12e Ave N, Sherbrooke, Qc J1H 5H3; e-mail: gerald.gahide@usherbrooke.ca; @ggahide

Indicates article with online supplemental data.

http://dx.doi.org/10.3174/ajnr.A7132

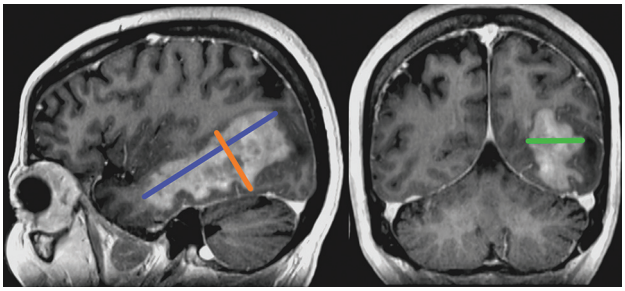

FIG 1. 3D measurement example on postcontrast 3D T1 MPRAGE. The *blue line* represents the longest 1D. The *orange line* is the longest length perpendicular to the *blue line*; their product represents 2D. The *green line* is the longest length perpendicular to the *blue* and *orange lines*; those 3 measures are used to calculate the volume for 3D.

Since the publication of the international guidelines in 2005, no clinical study has evaluated the agreement among the different measurement techniques for defining tumor-response categorization in PCNSL.[5] Hence, our study consisted of assessing the agreement among different measurement techniques in a group of patients treated at our institution for PCNSL. We also assessed whether a correlation was observed between response categorization and clinical surrogates.

MATERIALS AND METHODS

Patients

All immunocompetent patients presenting to our institution with PCNSL from 2000 to 2019 were considered for inclusion in this study. The inclusion criteria were as follows: 1) biopsy-confirmed PCNSL, 2) naïve patient with no prior treatment, 3) immunocompetency, and 4) 18 years of age or older. The exclusion criteria were as follows: 1) systemic lymphoma with secondary cerebral involvement, 2) relapsing disease, 3) unavailable imaging files, and 4) history of cytoreductive surgery.

Imaging Techniques

Cerebral MR imaging was performed before each cycle of treatment every 4–6 weeks. MR imaging was performed on a 1.5T scanner (Magnetom Symphony; Siemens) or a 3T scanner (Ingenia; Philips Healthcare). The imaging protocol included axial T1-weighted spin-echo, T2-weighted spin-echo, T2 FLAIR, diffusion-weighted imaging (multisection spin-echo single-shot echo-planar) with generated ADC maps and a T1-weighted gradient-echo acquisition after gadolinium injection (MPRAGE). Images were interpreted on a Barco MDNC 3421 reading station (Barco, Kortrijk, Belgium) with a PACS server.

Measurements

The measurements were obtained by 2 investigators (K.M.-T. and D.V.).

The axial 1D measurement was obtained by looking solely at the axial plane by adding the sum of the longest diameters of all enhancing lesions visible on axial images. Axial 1D measurements were expressed in millimeters.

The RECIST 1.1 measurement criteria consisted of the summation of the longest axial diameters for a maximum of 2 lesions per organ (lesions of ≥10 mm at baseline). Because the CNS equals 1 organ, 2 lesions at maximum were considered target lesions. The other lesions were considered nontarget but could influence the response categorization as recommended by the RECIST 1.1 guidelines.[6] In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm.

The longest 1D measurement was the sum of the longest diameters of all enhancing lesions among the measurements obtained on axial, coronal, or sagittal planes (Fig 1, blue line). In other words, the longest 1D measurements were made by adding the length of each lesion's longest axis. The longest 1D measurements were expressed in millimeters.

The 2D measurements corresponded to the sum of the products of the longest length of enhancing lesions with their maximum perpendicular diameter obtained in the same plane (axial, coronal, or sagittal) (Fig 1, orange line). 2D measures were expressed in square millimeters.

The 3D volumes were calculated using the 2 lengths obtained for calculating the 2D measurements and the longest perpendicular diameter (Fig 1, green line). The volume of each lesion was estimated by an ellipsoid formula [$V = (4 \times \pi \times A \times B \times C) / 3$] and expressed in cubic millimeters. If there was >1 lesion, 3D measurements corresponded to the summation of the different volumes.

As recommended by the guidelines, the size of a measurable lesion needed to be at least twice the thickness of the axial section acquisition and visible on ≥2 axial slices with 0-mm skip.[6] Considering the potential margin of error when measuring smaller lesions, a minimum of 1 cm in length in 1 axis was required to be considered measurable.[3,6] If an initially measurable lesion decreased to measure less than 5 mm during treatment, it was reported to measure 5 mm until complete resolution. As proposed in the modified Macdonald criteria and Küker et al,[2] a complete response (CR) was a residual enhancing lesion of <5 mm on 2 consecutive MR images in the absence of edema in the region of the biopsy, hemorrhage, or infection, or, if the initial lesion was >5 cm for 2 consecutive MR images.[7] Lesions of <1 cm or having nonnodular enhancement at baseline were reported as a nonmeasurable disease but were considered, and their evolution was noted as stable, increasing, or decreasing. If a nonmeasurable lesion grew and met measurable criteria on follow-up MR imaging, it was then properly measured.[6]

Interobserver variability was assessed by comparing the successive measurements of 19 lesions in 9 patients on 51 MR images. Intraobserver variability was assessed by repeating the measurements of 55 lesions on 20 baseline MR images for 20 patients after a 1-year delay.

Definition of Response Categories

Patients were classified into 4 groups according to the cutoff values reported in the Online Supplemental Data: PD, stable disease (SD), PR, and CR. The best overall response for each patient was considered to define the response category as recommended.[5,8] To assess the impact on response categorization of using the 25% and 50% cutoffs for unidimensional measurements, as recommended by the international guidelines for PCNSL, we performed a second analysis using these criteria proposed by Abrey et al[5] as reported in the Online Supplemental Data. For each

measurement technique, the lesions were compared with the baseline MR image and classified as CR, PR, SD, or PD according to the percentages indicated in the Online Supplemental Data. Any new lesion was considered a progressive disease except if a complete response had been obtained beforehand. In this case, the new lesion was considered a relapse. All complete responses had to be confirmed on follow-up imaging.

Table 1: Included patient characteristics (*N* = 40)

Patient Characteristics	Value (SD)
Mean age (yr)	61.5 (11.1)
Age range (yr)	29–81
Male/female ratio	23:17
Pathology	
Primary CNS large B-cell lymphoma	40
Prior WBRT	5
Median follow-up (yr)[a]	3.01
Total baseline MR imaging	33
Total baseline CT	7
Total follow-up MR imaging	254
Total follow-up CT	10
Mean number of MR images per patient	7.60 (3.49)
Mean number of lesions on baseline MR image	2.33 (1.14)
Disease location	
Superficial	16
Deep	15
Mixed	9
Particular disease distribution	
Midline crossing lesion	10
Multifocal, bilateral lesions	16
Induction response[a]	22
Relapse after induction[a]	12

Note:—WBRT indicates whole brain radiotherapy; SD, standard deviation.
[a]One patient was lost to follow-up, and 5 stopped their treatment for palliative care and were not included in this analysis.

Statistical Analysis

Statistical analysis was performed with SPSS software, Version 24 (IBM).

Interobserver and intraobserver correlations were evaluated by intraclass correlation coefficients (ICCs) with a 95% confidence interval. A single-measurement, absolute-agreement, 2-way mixed-effects model was used for calculation. Completely resolved lesions on follow-up MR images were excluded to not artificially increase inter- and intraobserver concordance.

Correlation between raw measurements according to the different methods was realized with the Spearman rank correlation after excluding resolved lesions, as aforementioned, to not artificially increase the coefficients. To determine the concordance among the different methods, we applied a cubic root to volumetric measurements and a square root to the surface area for comparison in the same unit of measurement (so-called root manipulation in the current article).

To assess the response categorization, we calculated the variation in size on follow-up MR images in comparison with the baseline MR image. The correlation coefficients among the different techniques were obtained after applying a cubic root on volumetric measurements and a square root on surface areas. These data were evaluated in pairs with the Spearman rank correlation (axial 1D versus longest 1D, axial 1D versus 2D, and so forth).

Response categories for each measurement technique were classified at each follow-up. The frequency of agreement was evaluated with a contingency table.

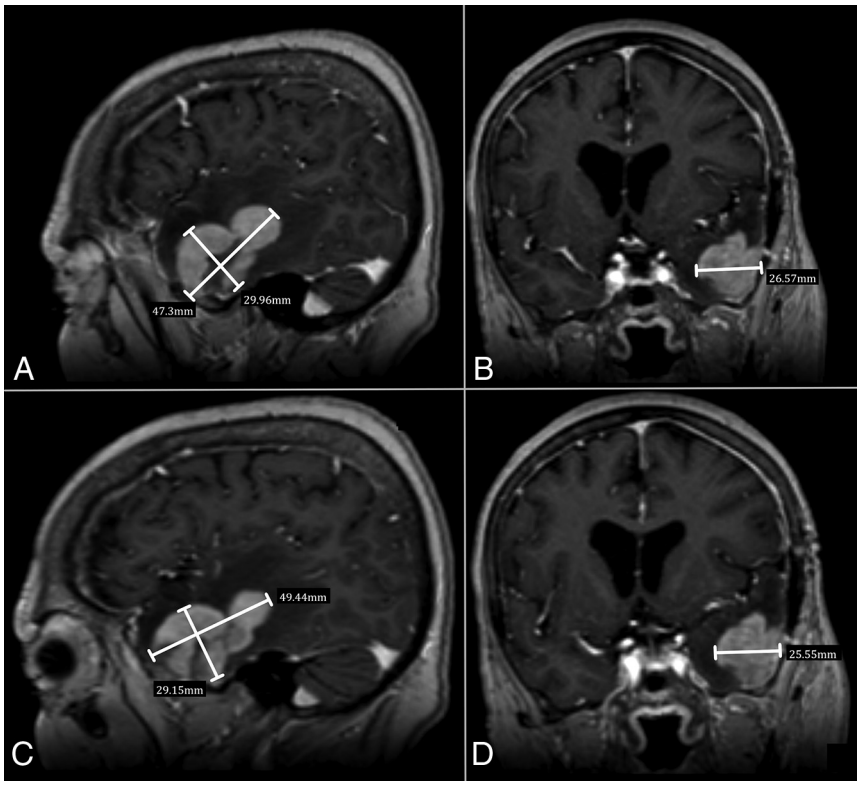

FIG 2. Postcontrast 3D T1 MPRAGE in a 56-year-old woman treated for a left temporal PCNSL. Comparison of the 3D measurements performed by the 2 readers (reader 1, *A* and *B*; reader 2, *C* and *D*).

RESULTS

Population Characteristics

An archive chart review identified 92 patients with a cerebral lymphoma diagnosis. From this population, 5 had a relapse of PCNSL and 17 had systemic lymphoma; 70 patients met the inclusion criteria. Nineteen patients were treated before 2003, and their MR images were not available in the data-storage archives. Eleven patients had a prior surgical resection instead of a biopsy, so 40 patients were included in the study. The mean age was 61.5 (SD, 11) years, with 42.5% women ($n = 17$) (Table 1). A total of 304 MR images were analyzed with a mean number of 7.6 MR images per patient (Fig 2).

Intraobserver and Interobserver Correlations

Both intra- and interobserver ICCs for 3D, 2D, and longest 1D measurements

Table 2: Intraobserver and interobserver correlations calculated with the ICC for each measurement method in raw units (95% CI)

Measurement Method	Intraobserver Correlation (95% CI)	Interobserver Correlation (95% CI)
Longest 1D	0.993 (0.985–0.997)	0.967 (0.947–0.979)
2D	0.997 (0.994–0.998)	0.982 (0.969–0.989)
3D	0.993 (0.985–0.997)	0.992 (0.986–0.995)

Table 3: Correlation between raw measurements with the Spearman rank correlation coefficient

Measurement Method	Spearman Rank Correlation Coefficient (95% CI)
Longest 1D vs axial 1D	0.977
Longest 1D vs RECIST	0.930
Longest 1D vs 2D	0.965
Longest 1D vs 3D	0.935
Axial 1D vs RECIST	0.915
Axial 1D vs 2D	0.963
Axial 1D vs 3D	0.942
RECIST vs 2D	0.924
RECIST vs 3D	0.898
2D vs 3D	0.990

showed very strong correlations (Table 2). The intraobserver ICC varied from 0.993 to 0.997 in raw units and was calculated to be 0.997 after root manipulation for 2D and 3D measurements. The interobserver ICC varied from 0.967 to 0.992 in raw units and from 0.966 to 0.968 after root manipulation for 2D and 3D, respectively.

Correlation among Measurement Strategies

The correlations among the different measurement techniques were also strong. If the tridimensional measurements were considered as the reference, the correlations with bidimensional and unidimensional measurements were also excellent, ranging from 0.99 (2D) to 0.90 (RECIST 1.1) (Table 3 and Fig 3).

A decrease in the correlation between RECIST 1.1 and the other techniques was noticed after the fifth MR image (Fig 3). From then on, fewer data were available because several patients were in complete remission or opted for palliative treatment and had thus been excluded from the analysis. As a reference, only 12 of the 40 patients remained at the sixth follow-up MR imaging.

Correlation among Different Response Measurements

Compared with the baseline MR images, most measurement methods presented with a correlation coefficient of ≥0.80 (Table 4).

2D had the best correlation with 3D, with a coefficient of 0.980, closely followed by the longest 1D versus 2D with a coefficient of 0.96. Compared with 3D, 2D (0.98) and the longest 1D (0.92) had a better correlation coefficient than axial 1D (0.83) and RECIST (0.79), which were still, nonetheless, excellent.

Response Categorization

Each patient was categorized as having CR, PR, SD, or PD according to the measurable disease extent evolution in comparison with the baseline. Agreement between each measurement method is presented as contingency tables (Online Supplemental Data). These results demonstrate excellent agreement in categorization between each technique, most being >90%, particularly for the first method.

DISCUSSION

Background, Reproducibility, and Measurements

One of the particular aspects of PCNSL is its radiologic presentation as a strongly enhanced lesion.[2] This aspect makes PCNSL quite simple to delineate and measure, very likely explaining the excellent inter- and intraobserver correlations reported herein (both being >0.96). Nevertheless, PCNSL may behave oddly, growing asymmetrically in several directions and giving a nonspheric aspect to the lesion. This feature implies that there could be some variations in estimating the lesion behavior according to the kind of measurement used (1D, 2D, or 3D).

The response assessment of PCNSL is currently based on the widely accepted international guidelines for standardizing evaluation and response criteria. These guidelines have dramatically improved the quality of patient follow-up and management and have participated in the improvement of scientific publications on the subject, allowing reliable comparisons among studies. One of the interesting points of these guidelines was the decision of the authors not to choose between recommending unidimensional or bidimensional measurements.[5,9,10] This decision was very likely due to the different backgrounds of the authors: Some coming from the 2D neuro-oncologic world used to the modified Macdonald criteria, and the others coming from the 1D solid-tumor world used to the RECIST.

Most interesting, our study demonstrated excellent correlation coefficients among all the different techniques for evaluating the variation in the size of the lesions compared with the baseline MR image. The longest 1D and 2D had excellent correlations with the 3D reference, ranging from 0.92 to 0.94 for the longest 1D and an almost perfect correlation of 0.98 to 0.99 for 2D. This finding could slightly advocate using 2D measurements, given their almost perfect correlations with 3D. Nevertheless, all the other measurement techniques had correlation coefficients greater than acceptable, and, in routine practice, the simplest measurement technique needs particular consideration because it is generally more acceptable for readers and accessible to all clinicians. Many contributing factors to discrepancies have long been recognized, most particularly excessive workload and cognitive overload, particularly in oncologic centers with heavy radiologic CT and MR imaging activities.[11]

One of the other interesting points of these international guidelines was the choice of the authors to recommend 25% and 50% cutoffs to define progressive disease and partial response for both unidimensional and bidimensional measurements. As indicated by Therasse et al[12] in the RECIST 1.0 guidelines, a 25% increase in size for bidimensional measurements corresponds to a 12% increase for unidimensional measures. Due to observer reproducibility issues, a 12% limit was found to be prone to mistakes, and a cutoff of 20% was chosen for the RECIST 1.0 criteria. They also indicated that a 50% decrease in size for bidimensional measurements corresponds to a 30% decrease for unidimensional measurements. Our study tends to confirm that using 25% and 50% cutoffs for unidimensional measurements provides less homogeneous response categorization among the different techniques. This is not readily obvious when just reading the excellent correlation coefficients obtained in both cases, but those coefficients are heavily influenced by

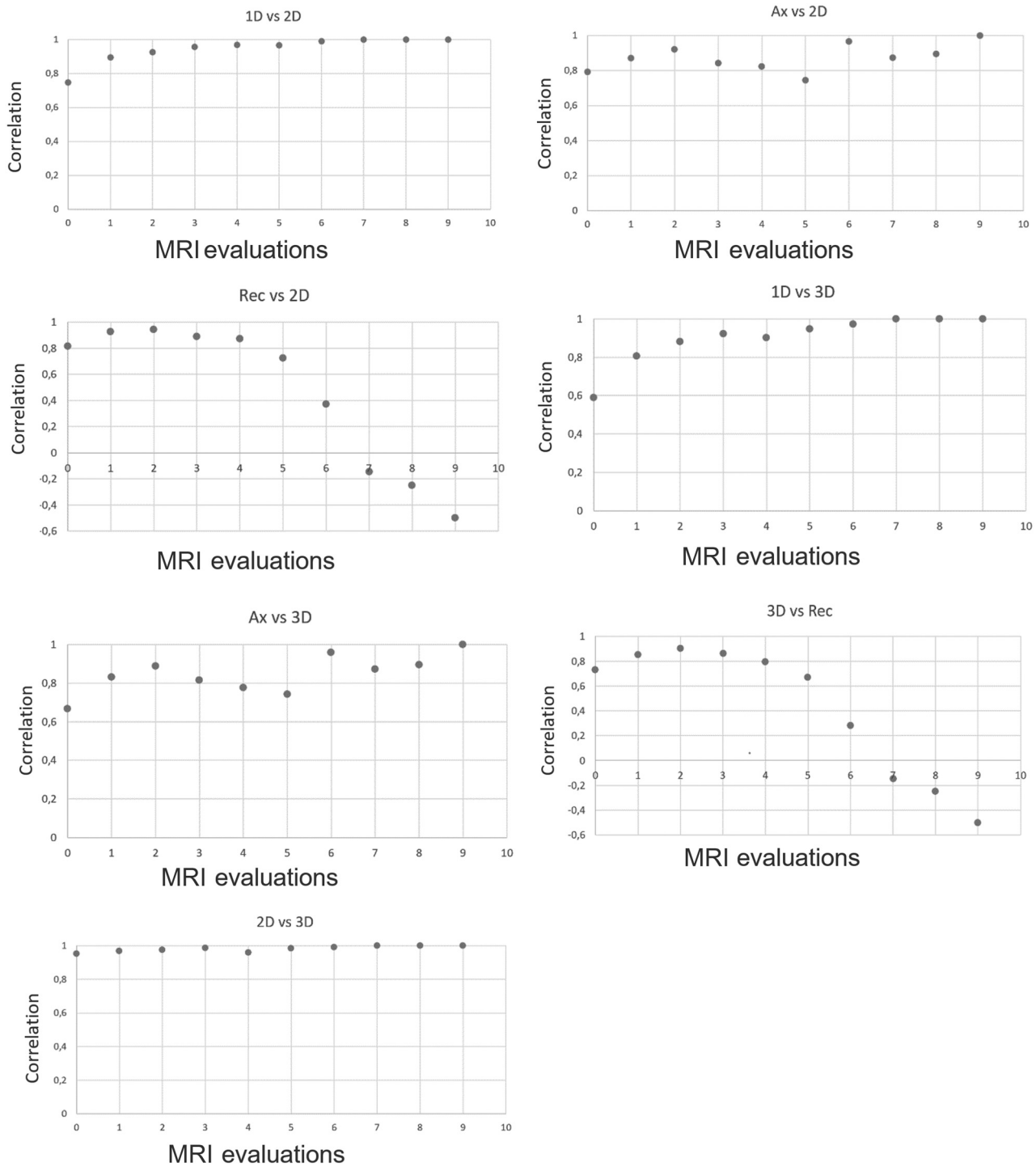

FIG 3. Correlation of raw measurements calculated with the Spearman correlation after we applied cubic root on volume calculated with the 3D method and square root on surface area calculated with the 2D methods (exclusion of zero values). Rec indicates RECIST; Ax, axial.

the complete responses encountered for >50% of the patients in our series. These complete responses have perfect agreement, which is obvious because they correspond to the complete disappearance of tumoral enhancement. On the other hand, there was more dispersion for the other categories when using 25% and 50% for unidimensional measurements. Stable disease ranged from 3 to 7 patients, and progressive disease ranged from 6 to 10 with 25% and 50% cutoffs, while ranges were from 4 to 5 patients and 6 to 9 with 20% and 30% cutoffs, respectively. Similarly, the frequencies of agreement were all stronger or equal when using 20% and 30% cutoffs rather than 25% and 50%. This finding advocates using 20% and 30% cutoffs for unidimensional measurements instead of the recommended 25% and 50% cutoffs by Abrey et al.[5]

Axial 1D and longest 1D could both be used as standard measurements for PCNSL, because they showed great correlation

Table 4: Spearman rank correlation coefficients of the lesion variation in size compared with baseline MR images[a]

Measurement Method	Spearman Correlation Coefficient (95% CI)
Longest 1D vs axial 1D	0.890
Longest 1D vs RECIST	0.898
Longest 1D vs 2D	0.962
Longest 1D vs 3D	0.918
Axial 1D vs RECIST	0.796
Axial 1D vs 2D	0.872
Axial 1D vs 3D	0.831
RECIST vs 2D	0.847
RECIST vs 3D	0.794
2D vs 3D	0.980

[a] Cubic roots and square roots were applied for 3D and 2D measurements, respectively.

compared with 3D measurements and had an excellent response categorization agreement. However, the longest 1D showed a better performance in all our analyses. According to the literature, 1D, 2D, and 3D measurements are equivalent to volumetric criteria and are easier to perform.[8,13] Many studies have already demonstrated great intraobserver and interobserver reliability using unidimensional measurements.[14,15] Most guidelines suggest unidimensional measurements in oncologic follow-up to assure greater reproducibility and to facilitate follow-up. Our results show that unidimensional measurements could also be applied to PCNSL.

Limitations

Using the Spearman correlation, the multiplicity of follow-up MR images could induce an overestimation of the correlations obtained with raw measures. This could have affected the results presented in Table 3. However, the presentation of the same results in Fig 2 also expresses this correlation without the potential error of overestimation. Because Table 3 is easier to read, we decided to present both methods.

CONCLUSIONS

Unidimensional and bidimensional measurements are both reliable techniques to assess the PCNSL response to treatment, though there was a slight advantage of 2D measurements regarding the correlation coefficients obtained in comparison with 1D and 3D measurements and for classifying the response categories. Our study suggests that the longest 1D measurements could be used for the follow-up of PCNSL with high performance and agreement with 3D measurements. Our study also indicated that if unidimensional measurements were to be used, 20% and 30% cutoffs should be used instead of 25% and 50% for defining PD and PR, respectively, contrary to the recommendations of the international guidelines.

ACKNOWLEDGMENTS

We wish to thank the biostatistics team from Centre Hospitalier Universitaire de Sherbrooke Research Center, and Dr M. St-Amant Beaudoin for his advice.

Disclosures: Gerald Gahide—*UNRELATED*: *Consultancy:* Boston Scientific.

REFERENCES

1. Miller AB, Hoogstraten B, Staquet M, et al. **Reporting results of cancer treatment.** *Cancer* 1981;47:207–14 CrossRef Medline
2. Küker W, Nägele T, Thiel E, et al. **Primary central nervous system lymphomas (PCNSL): MRI response criteria revised.** *Neurology* 2005;65:1129–31 CrossRef Medline
3. Lin NU, Lee EQ, Aoyama H, et al. Response Assessment in Neuro-Oncology (RANO) group. **Response assessment criteria for brain metastases: proposal from the RANO group.** *Lancet Oncol* 2015;16:270–78 CrossRef Medline
4. Vogelbaum MA, Jost S, Aghi MK, et al. **Application of novel response/progression measures for surgically delivered therapies for gliomas: Response Assessment in Neuro-Oncology (RANO) working group.** *Neurosurgery* 2012;70:234–43 CrossRef Medline
5. Abrey LE, Batchelor TT, Ferreri AJ, et al. International Primary CNS Lymphoma Collaborative Group. **Report of an international workshop to standardize baseline evaluation and response criteria for primary CNS lymphoma.** *J Clin Oncol* 2005;23:5034–43 CrossRef Medline
6. 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;45:228–47 CrossRef Medline
7. Macdonald DR, Cascino TL, Schold SC, et al. **Response criteria for phase II studies of supratentorial malignant glioma.** *J Clin Oncol* 1990;8:1277–80 CrossRef Medline
8. Galanis E, Buckner JC, Maurer MJ, et al. **Validation of neuroradiologic response assessment in gliomas: measurement by RECIST, two-dimensional, computer-assisted tumor area, and computer-assisted tumor volume methods.** *Neuro Oncol* 2006;8:156–65 CrossRef Medline
9. Chiavazza C, Pellerino A, Ferrio F, et al. **Primary CNS lymphomas: challenges in diagnosis and monitoring.** *Biomed Res Int* 2018;2018:3606970 CrossRef Medline
10. Grommes C, Rubenstein JL, DeAngelis LM, et al. **Comprehensive approach to diagnosis and treatment of newly diagnosed primary CNS lymphoma.** *Neuro Oncol* 2019;21:296–305 CrossRef Medline
11. Vendrell JF, Frandon J, Boussat B, et al. **Double reading of outsourced CT/MR radiology reports: retrospective analysis.** *J Patient Saf* 2018 Aug 13. [Epub ahead of print] CrossRef Medline
12. Therasse P, Arbuck SG, Eisenhauer EA, et al. **New guidelines to evaluate the response to treatment in solid tumors.** *J Natl Cancer Inst* 2000;92:205–16 CrossRef Medline
13. Shah GD, Kesari S, Xu R, et al. **Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas.** *Neuro Oncol* 2006;8:38–46 CrossRef Medline
14. Hopper KD, Kasales CJ, Van Slyke MA, et al. **Analysis of interobserver and intraobserver variability in CT tumor measurements.** *AJR Am J Roentgenol* 1996;167:851–54 CrossRef Medline
15. McErlean A, Panicek DM, Zabor EC, et al. **Intra- and interobserver variability in CT measurements in oncology.** *Radiology* 2013;269:451–59 CrossRef Medline
16. Chappell R, Miranpuri SS, Mehta MP. **Dimension in defining tumor response.** *J Clin Oncol* 1998;16:1234 CrossRef Medline