Received:
1 February 2016

Revised:
13 June 2016

Accepted:
21 July 2016

<http://dx.doi.org/10.1259/bjr.20160110>

Cite this article as:

Morris KA, Parry A, Pretorius PM. Comparing the sensitivity of linear and volumetric MRI measurements to detect changes in the size of vestibular schwannomas in patients with neurofibromatosis type 2 on bevacizumab treatment. *Br J Radiol* 2016; **89**: 20160110.

FULL PAPER

Comparing the sensitivity of linear and volumetric MRI measurements to detect changes in the size of vestibular schwannomas in patients with neurofibromatosis type 2 on bevacizumab treatment

^{1,2}KATRINA A MORRIS, MBBS (Hons), FRACP, ²ALLYSON PARRY, MB ChB (Hons), FRCP and ³PIETER M PRETORIUS, MB ChB, FRCR

¹St Vincent's Clinical School, University of New South Wales, Sydney, NSW, Australia

²Department of Neurology, John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

³Department of Neuroradiology, John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Address correspondence to: Dr Pieter Michiel Pretorius

E-mail: pieter.pretorius@ouh.nhs.uk

Objective: To compare the sensitivity of linear and volumetric measurements on MRI in detecting schwannoma progression in patients with neurofibromatosis type 2 on bevacizumab treatment as well as the extent to which this depends on the size of the tumour.

Methods: We compared retrospectively, changes in linear tumour dimensions at a range of thresholds to volumetric tumour measurements performed using Brainlab iPlan® software (Feldkirchen, Germany) and classified for tumour progression according to the Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) criteria.

Results: Assessment of 61 schwannomas in 46 patients with a median follow-up of 20 months (range 3–43 months) was performed. There was a mean of 7 time points per tumour (range 2–12 time points). Using the volumetric REiNS criteria as the gold

standard, a sensitivity of 86% was achieved for linear measurement using a 2-mm threshold to define progression.

Conclusion: We propose that a change in linear measurement by 2 mm (particularly in tumours with starting diameters 20–30 mm, the majority of this cohort) could be used as a filter to identify cases of possible progression requiring volumetric analysis. This pragmatic approach can be used if stabilization of a previously growing schwannoma is sufficient for a patient to continue treatment in such a circumstance.

Advances in knowledge: We demonstrate the real-world limitations of linear vs volumetric measurement in tumour response assessment and identify limited circumstances where linear measurements can be used to determine which patients require the more resource-intensive volumetric measurements.

INTRODUCTION

The use of bevacizumab to control the growth of vestibular schwannomas (VSs) and to improve hearing outcomes in patients with neurofibromatosis type 2 (NF2) has rapidly expanded since initial reports 6 years ago.^{1–6} Radiological response assessment is an important part of the follow-up of these patients, and the Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) consortium has recommended volumetric tumour measurements for clinical trials with a 20% change in volume relative to baseline defined as the criteria for radiological response.⁷

Volumetric measurement is sensitive for monitoring changes in tumour size, particularly the complex multi-lobulated lesions observed in NF2 where multiple collision

lesions can contribute to mass effect.^{7,8} Harris et al⁹ found a three-dimensional (3D) volumetric technique to be significantly more sensitive than the cube of a single dimension in identifying tumour progression, although they recognized the extra resources required for the 3D volumetric methodology.

Although volumetric measurements are sensitive to relatively small changes in schwannoma size, this technique has its own limitations at a practical level. To measure schwannomas reliably, a semi-automated technique is required with checking of regions of interest by a trained observer. It has been estimated that at least five slices and slice thickness <3 mm through a lesion must be used for accurate volume calculation,¹⁰ thus limiting the

measurement of small tumours; in addition, specialist software and the time of either a specialist radiologist or trained technician is required above that required for standard clinical reporting. We have calculated that the estimated additional time to segment a VS for volumetric assessment in this patient group is approximately 20 min, with the precise time required being dependent on the complexity of the lesion. The location of VSs provides difficulties for automated segmentation, especially in the context of collision lesions or adjacent meningiomas.

These extra time and cost implications may make volumetric measurements difficult to implement in some centres. Funding the extra time and IT resources required for a radiologist or a technician to perform volumetric measurements in addition to usual clinical reporting is not currently achievable in the majority of centres in the UK. The need for timely provision of schwannoma response data is essential to clinical care, but the resource implications makes this difficult to achieve outside of a funded clinical trial. This raises the question of whether less resource-intensive linear tumour measurements retain a role in tumour response assessment in patients with NF2 on bevacizumab treatment.

In this retrospective study, we compare the utility of a 3D semi-automated volumetric measurement technique against the standard linear measurements that are routinely performed on these patients' MRI scans. The motivation for this study is to find a "real-world" model of care that balances the need for careful monitoring of patients for treatment failure (demonstrated by ongoing schwannoma growth) and the time and expense limitations of volume measurement. As a minimum, we hope to understand any potential limitations from reliance on linear measurements alone.

When considering the results of this study, it is important to appreciate that the main purpose of our MRI surveillance is to confirm that the index schwannoma requiring treatment has not increased in size despite treatment with bevacizumab (one of the primary aims of treatment being to stabilize or reduce the size of a previously growing schwannoma).

Our hypothesis is that linear measurements are sufficiently sensitive to exclude significant schwannoma progression in the majority of patients, and that in practice, they may be used to identify the smaller group of patients where volumetric measurements are subsequently needed to confirm the REiNS definition of tumour progression. By using this approach, we hypothesize that the number of lesions requiring volumetric assessment (with the attendant resource requirements) may be reduced during follow-up.

METHODS AND MATERIALS

The MRI scans of patients with NF2 treated with bevacizumab for growing schwannomas as part of a UK prospective observation study⁶ were serially measured using both volumetric and linear measurements. The tumours being treated were either VS or lower cranial nerve schwannomas. Schwannomas with a volume $<1\text{ cm}^3$ were excluded from analysis due to the reduced

accuracy associated with the measurement of small tumour volumes.^{9,10}

Volume measurements

Measurements were performed using Brainlab iPlan® (Feldkirchen, Germany), a semi-automated software package, by a single trained operator under the supervision of a neuroradiologist. Tumour contours were manually checked and corrected as required on each slice. The neuroradiologist confirmed consistent lesion contours had been performed by the trained operator who performed volumetric contours blinded to the trend in linear measurements. Volumetric assessments were performed for each patient using either a gadolinium-enhanced T_1 weighted sequence with slice thickness of $\leq 3\text{ mm}$ or a $\leq 1\text{-mm}$ high-resolution heavily T_2 weighted steady-state sequence. These sequences (FIESTA, CISS or balanced-FFE depending on the scanner manufacturer) were also acquired after gadolinium enhancement to take advantage of the contrast enhancement seen on these images that enable better delineation of the margins of larger tumours.¹¹

Scans from multiple time points were assessed at a single session for a given patient to ensure consistency of inclusion of elements of a collision lesion if present. For any given patient, the same sequence was used throughout radiological surveillance.

Radiological responses were defined according to the REiNS criteria: partial response was a $\geq 20\%$ reduction in tumour volume, tumour progression as $\geq 20\%$ increase in tumour volume and stable disease as growth or reduction in tumour volume of $<20\%$ when compared with baseline tumour volume.

Images were excluded from volumetric assessment if: there was artefact from an auditory brainstem or cochlear implant preventing accurate volumetric assessment; imaging was unsuitable for consistent serial volumetric measurement; or tumour volume was $<1\text{ cm}^3$.

Coefficient of variance of volumetric measures

To calculate the intrarater reproducibility, blinded to previous segmentation, a tumour at a subset of time points was segmented and measured as described above. A time point from the series which was not being replicated was reviewed with regard to any assumptions or rules with regard to handling collision lesions. The coefficient of variance was calculated on repeated values. Tumours were selected at random for inclusion in the coefficient of variance (CoV) analysis.

Linear measurements

The MRI scans of patients treated with bevacizumab for NF2-associated schwannomas in the UK were reported by the same experienced neuroradiologist who also supervised the volumetric measurement analysis. However, these measurements were made independent of and prior to the volumetric measurement sessions. As standard practice for reporting, scans at multiple time points were reviewed and the sequence and slice number was documented for each time point to ensure that the same sequence was used over the period of monitoring and that the same components of the tumour were included in the

measurements to allow consistency of measurement between scans. Axial images were always used and up to three standardized measurements were performed on axial images, whereas one coronal measurement (height) was performed when an appropriate coronal sequence was available. The measurements always included the cisternal portion of the tumours. A previous unpublished audit of the neuroradiologist performing these measurements had a variation of no more than 2 mm in any dimension on serial measurements of VSs.

Changes over time in the linear dimensions of the schwannomas were reviewed to designate a change in schwannoma size into the appropriate outcome category.

Two different thresholds were applied to the reported single-dimensional (1D) linear measurements to categorize whether the schwannoma had exhibited tumour response or tumour progression when compared with the pre-treatment linear value (with intermediate changes in tumour size falling into the category of tumour stability):

- (i) a change of +2 mm (progression) or −2 mm (response) in *any* consistent solitary serial linear measure of the tumour (1D 2 mm)
- (ii) a change of +3 mm (progression) or −3 mm (response) in *any* consistent solitary serial linear measure of the tumour (1D 3 mm).

A two-dimensional (2D) *combined* linear assessment was also made by summing the change in width and depth measurements of the cerebellopontine angle component of the schwannoma or of the lower cranial nerve schwannoma. Three different thresholds were applied to this 2D combined linear assessment:

- (i) a change of >10% (2D 10%)
- (ii) a change of >20% (2D 20%)
- (iii) a change of 3 mm or more (2D 3 mm).

Comparison of methods

The sensitivity and specificity of the five linear measurement thresholds defined above to detect schwannoma progression and response were compared against the REiNS criteria using the semi-automated volumetric technique.

Statistical analyses

Statistical analyses were performed using GraphPad Prism® v. 6.0f (GraphPad Software Inc., La Jolla, CA). Log-rank tests were used to compare the Kaplan–Meier curves of the volumetric and linear methods with a *p*-value of <0.01 taken as significant to take into account multiple comparisons.

RESULTS

61 schwannomas in 46 patients (59 VSs and 2 lower cranial nerve schwannomas) were evaluated. There was a mean of 7 time points per tumour (range 2–12 time points). The median duration of follow-up from the pre-treatment baseline MRI scan to the final MRI scan was 20 months (range 3–43 months). The median schwannoma volume at baseline was 5.63 cm³ (range 1.04–58.63 cm³). Schwannoma diameter was >10 mm when the volume was >1 cm³.

Figure 1. The relationship between coefficient of variance and tumour volume.

Replicate volume measurements were made on 25 tumour time points to calculate CoV. The median CoV for semi-automated volumetric schwannoma measurement of all schwannoma volumes (independent of size) was 3% (range 0.02–11.95%) (Figure 1). The median coefficient of variance improved to 2.4% (range 0.02–6.93%) when only schwannoma volumes >1 cm³ were included in the variance analysis.

Identification of ongoing schwannoma growth (progression) or regression (response)

The number of schwannomas meeting the criteria for either tumour progression or tumour response for each of the five linear measurement thresholds and also when defined by volumetric measurement is shown in Table 1. If a linear measurement was used as a filter to select cases for subsequent volumetric measurement, the number of schwannomas requiring volumetric measurements to identify progressing tumours would be reduced from 61 to 25 for the 1D 2-mm filter and to 18 for the 2D 3-mm filter.

There was a significant difference (*p* = 0.004) in the number of schwannomas that met the criteria for progression when the 2D 20% threshold was used, when compared with all other response thresholds. There was no significant difference between the Kaplan–Meier curves for the other linear thresholding criteria (1D 2 mm, 1D 3 mm, 2D 3 mm and 2D 10%) and the number of

Table 1. Tumours defined as progression or response by each measurement technique

Measurement technique	Number of schwannomas that progressed	Number of schwannomas that responded
Volumetric	14	24
1D 2 mm	25	42
1D 3 mm	21	25
2D 3 mm	18	34
2D 10%	8	10
2D 20%	4	8

1D, single dimension; 2D, two dimensions.

Figure 2. Kaplan-Meier analysis of time to schwannoma progression or censor. 1D, single dimension; 2D, two dimensions.

schwannomas meeting the REiNS criteria for tumour progression (by volumetric assessment) (Figure 2).

The sensitivity and specificity of each of the five linear measurement methods to determine schwannoma progression or response when compared with using a 20% change in schwannoma volume is shown in Table 2. Awaiting a confirmatory linear measure on a second scan 3 months later did not significantly improve the diagnostic utility of the linear measurements.

The cases identified as having met criteria for progression using volumetric and linear methods, respectively, were not identical. Two cases identified on volume measures to have met the REiNS criteria and grown by >20% were not identified as having met the criteria for progression using any of the five linear measures. The volumes of these two schwannomas were 1.0 and 1.6 cm³, respectively. Progression on volume measures was identified at final follow-up in one of these cases and at the 6-month time point prior to the final follow-up scan in the second case.

The median (range) time delay for the detection of tumour progression using the 1D 2 mm, 1D 3 mm, 2D 10% and 2D 3 mm methods when compared with volumetric analysis was 4 (–10 to 18), 7 (0 to 18), 8 (–16 to 28) and 4 (–13 to 27) months. There was a large range of values for the time to detection of progression in all linear measures, particularly when

using 2D 3 mm, although this difference was not statistically significant compared with 1D 2 mm ($p = 0.9$).

Taking 1D 2 mm as a possible candidate threshold in practice compared with volumetric measurements, an exploratory analysis of the effect of schwannoma diameter on sensitivity and specificity was undertaken (Table 3). No schwannomas >30 mm diameter showed progression according to the REiNS criteria, and thus sensitivity cannot be calculated for these groups.

DISCUSSION

The advent of non-surgical treatments for NF2-associated tumours requires the use of robust outcome tools to monitor treatment response and guide the clinical management of the patient.

Volumetric measurement is the gold standard for monitoring tumour response to bevacizumab in patients with NF2. Given the treatment protocol, detecting tumour progression is of greater clinical importance than detecting either stable disease or disease response. The time and resource intensive nature of volumetric measurement means that linear measurements are still the more commonly performed measurement technique in many centres.

The geometry of roughly spherical objects (such as VSs) means that any change in diameter corresponds to a disproportionately large change in volume, *e.g.* a doubling in diameter corresponds to an eight-fold increase in volume. It should therefore be clear that a modest change in volume (such as the 20% increase defined as tumour progression) would correspond to a relatively small change in linear (diameter) measurements. The limited resolution of MRI scans—and the inherent measurement errors that make linear changes of <2 mm difficult to detect reliably—represent a natural barrier to detecting a linear change corresponding to a 20% volume change in smaller tumours. This is in keeping with our finding that sensitivity of a 1D 2-mm change in diameter is greater for tumours of 21–30 mm than smaller tumours. It is reasonable to expect the sensitivity of this threshold to increase further for tumours >30 mm, but unfortunately, none of the tumours in that size range showed progression in our cohort according to the REiNS criteria, and the performance of the linear measurement could therefore not be assessed.

Table 2. Sensitivity and specificity of linear thresholds compared with the Response Evaluation in Neurofibromatosis and Schwannomatosis criteria

Linear threshold	Tumour progression		Tumour response	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
1D 2 mm	86	72	96	49
1D 3 mm	71	77	79	84
2D 3 mm	86	89	79	59
2D 10%	57	98	42	86
2D 20%	29	100	17	89

1D, single dimension; 2D, two dimensions.

Table 3. Sensitivity and specificity of single dimension (1D) 2 mm by tumour diameter

Diameter (mm)	Sensitivity (%)	Specificity (%)	Number ^a
10–20	75	100	10
21–30	90	73	32
31–40	^b	75	16
41–50	^b	63	8
>50	^b	50	2

^aFive tumours had diameters in two different categories that could be used for assessment over time.

^bSensitivity cannot be calculated due to no true positive events in >30-mm schwannomas.

The coefficient of variance of volumetric measurements in this study was in keeping with those previously reported and below the estimated variance of 5–10% that was used to calculate the 20% volumetric response and progression criteria set out by REiNS.^{7,9,10} Although linear measures were not as sensitive or specific as volumetric measures for the identification of tumour progression, they too are highly replicable if performed by a single experienced neuroradiologist.

It is possible that identifying a change in volume of a small lesion might have less significant impact on clinical decision making. A change in treatment strategy for a smaller lesion may not need to be considered as acutely as in the case of a larger lesion if causing neurological compromise and approaching the size limits for optimal surgical or radiosurgical management.

There is no significant difference between the sensitivity for detecting progression between 1D 2 mm and 2D 3 mm. There is a higher false-positive rate with 1D 2 mm (decreasing specificity); however, there is no extra calculation of a 2D measure required reducing the risk of error. In addition, there was no time delay benefit to using the 2D measure, in fact there was a non-significant decrease time variance of progression detection to volumetric measurements with the 1D 2-mm measurement.

Our study suggests that the use of linear measurements is more reliable in larger tumours and may obviate the need for

volumetric measurements in some larger tumours. However, we acknowledge that this is at the risk of either a delay in or the failure of identification of tumour progression, particularly in smaller tumours. In addition, this study has shown that the use of linear measurements may overestimate response relative to volumetric measures where a strict 20% decrease in schwannoma volume is required. However, importantly, there were no schwannomas mistaken as response by linear measurement criteria that had in fact progressed by volumetric criteria.

In conclusion, to define response in keeping with those described in the literature requires volumetric measurements. However, linear measurements can provide helpful insights as to schwannoma behaviour to guide management if there may be a delay in obtaining volumetric measurements. We propose that the use of linear measurements with a 2-mm change threshold as a filter to identify cases of possible progression requiring volumetric analysis could be a pragmatic approach that reduces the volumetric measurement burden. This approach can be used if stabilization of a previously growing schwannoma is sufficient for a patient to continue treatment in such a circumstance.

FUNDING

Dr Morris gratefully acknowledges the support of a Biomedical Research Centre–National Institute for Health Research grant from the University of Oxford.

REFERENCES

- Plotkin SR, Stemmer-Rachamimov AO, Barker FG 2nd, Halpin C, Padera TP, Tyrrell A, et al. Hearing improvement after bevacizumab in patients with neurofibromatosis type 2. *N Engl J Med* 2009; **361**: 358–67. doi: <http://dx.doi.org/10.1056/NEJMoa0902579>
- Mautner VF, Nguyen R, Kutta H. Bevacizumab induces regression of vestibular schwannomas in patients with neurofibromatosis type 2. *Neurooncology* 2010; **12**: 14–18. doi: <http://dx.doi.org/10.1093/neuonc/nop010>
- Alanin MC, Klausen C, Caye-Thomasen P, Thomsen C, Fugleholm K, Poulsen L, et al. The effect of bevacizumab on vestibular schwannoma tumour size and hearing in patients with neurofibromatosis type 2. *Eur Arch Otorhinolaryngol* 2015; **272**: 3627–33. doi: <http://dx.doi.org/10.1007/s00405-014-3398-3>
- Hochart A, Gaillard V, Baroncini M. Bevacizumab decreases vestibular schwannomas growth rate in children and teenagers with neurofibromatosis type 2. *J Neurooncol* 2015; **124**: 229–36. doi: <http://dx.doi.org/10.1007/s11060-015-1828-8>
- Blakeley JO, et al. (2016). Efficacy and Biomarker Study of Bevacizumab for Hearing Loss Resulting From Neurofibromatosis Type 2-Associated Vestibular Schwannomas. *J Clin Oncol* **34**(14): 1669–1675. doi: [10.1200/JCO.2015.64.3817](http://dx.doi.org/10.1200/JCO.2015.64.3817)
- Morris KA, Golding JF, Axon PR. Bevacizumab in neurofibromatosis type 2 (NF2) related vestibular schwannomas: a nationally coordinated approach to delivery and prospective evaluation. *Neurooncol Pract* 2016: 1–9.

- doi: <http://dx.doi.org/10.1093/nop/npv065>
7. Dombi E, Ardern-Holmes SL, Babovic-Vuksanovic D. Recommendations for imaging tumor response in neurofibromatosis clinical trials. *Neurology* 2013; **81**(Suppl. 1): S33–40. doi: <http://dx.doi.org/10.1212/01.wnl.0000435744.57038.af>
 8. Stivaros SM, Stemmer-Rachamimov AO, Alston R, Plotkin SR, Nadol JB, Quesnel A, et al. Multiple synchronous sites of origin of vestibular schwannomas in neurofibromatosis type 2. *J Med Genet* 2015; **52**: 557–62. doi: <http://dx.doi.org/10.1136/jmedgenet-2015-103050>
 9. Harris GJ, Plotkin SR, Maccollin M. Three-dimensional volumetrics for tracking vestibular schwannoma growth in neurofibromatosis type II. *Neurosurgery* 2008; **62**: 1314–19; discussion 1319–20. doi: <http://dx.doi.org/10.1227/01.neu.0000333303.79931.83>
 10. Snell JW, Sheehan J, Stroila M, Steiner L. Assessment of imaging studies used with radiosurgery: a volumetric algorithm and an estimation of its error. Technical note. *J Neurosurg* 2006; **104**: 157–62. doi: <http://dx.doi.org/10.3171/jns.2006.104.1.157>
 11. Shigematsu Y, Korogi Y, Hirai T, Okuda T, Ikushima I, Sugahara T, et al. Contrast-enhanced CISS MRI of vestibular schwannomas: phantom and clinical studies. *J Comput Assist Tomogr* 1999; **23**: 224–31. doi: <http://dx.doi.org/10.1097/00004728-199903000-00010>