

Inter-observer variability between radiologists reporting on cerebellopontine angle tumours on magnetic resonance imaging

S R TEH¹, S RANGUIS², P FAGAN²

¹University of New South Wales, Sydney, and ²Department of Otolaryngology, St Vincent's Private Hospital, Sydney, Australia

Abstract

Background: Studies demonstrate the significance of intra- and inter-observer variability when measuring cerebellopontine angle tumours on magnetic resonance imaging, with measured differences as high as 2 mm.

Objective: To determine intra- and inter-observer measurement variability of cerebellopontine angle tumours in a specialised institution.

Methods: The magnetic resonance imaging maximal diameter of 12 randomly selected cerebellopontine angle tumours were independently measured by 4 neuroradiologists at a tertiary referral centre using a standard definition for maximal tumour diameter. Average deviation and intraclass correlation were subsequently calculated.

Results: Inter-observer difference averaged 0.33 ± 0.04 mm (range, 0.0–0.8 mm). Intra-observer measurements were more consistent than inter-observer measurements, with differences averaging 0.17 mm (95 per cent confidence interval = 0.27–0.06, $p = 0.002$). Inter-observer reliability was 0.99 (95 per cent confidence interval = 0.97–0.99), suggesting high reliability between the readings.

Conclusion: The use of a standard definition for maximal tumour volume provided high reliability amongst radiologists' readings. To avoid oversizing tumours, it is recommended that conservative monitoring be conducted by the same institution with thin slice magnetic resonance imaging scans.

Key words: Cerebellopontine Angle; Neuroma; Acoustics; Observer Variation; Magnetic Resonance Imaging; Reproducibility Of Results

Introduction

The advent of readily available magnetic resonance imaging (MRI) has led to the knowledge that many relatively asymptomatic acoustic tumours show little growth. Many centres now adopt an expectant policy in these circumstances, if the tumour is less than 15 mm in the cerebellopontine angle. Serial MRI is the only intervention that is taken.

The senior author (PF) has noted that a great many patients express concern when told that a tumour is showing growth. However, it is necessary to ask what significant growth is and how important inter-observer reliability is, particularly when serial studies are carried out in different institutions by various radiologists utilising different technology. Furthermore, other studies have shown significant inter-observer differences in tumour measurement.^{1–5}

This study examined inter-observer reliability in a specialised institution.

Materials and methods

The MRI maximum diameter of 12 cerebellopontine angle tumours was independently measured by 4 experienced radiologists. The 12 cases were randomly chosen from a selection of patients who had received no treatment.

The MRIs were conducted using either a 3 T Siemens Magnetom Verio or a 1.5 T Siemens Magnetom Aera scanner (Munich, Germany). All sequences were performed with the three-dimensional sampling perfection with application-optimised contrast using different flip angle evolution ('SPACE') sequence protocol.

The images were examined using the virtual calipers on an Agfa Impax picture archiving and communication system (Mortsel, Belgium). To avoid bias, all radiologists cleared records of image analysis following each measurement on the imaging database.

The intervals measured included the largest dimension of the cerebellopontine angle moiety of the tumour in the

axial plane (in millimetres). In addition, ipsilateral and contralateral middle cerebellar peduncles were measured at the same level for comparison. This measurement was taken from the point of tumour–peduncle contact to the most lateral tip of the fourth ventricle. Inter-observer variability was determined between two separate measurements of the same case taken by the same radiologist after at least a 1-day interval.

The average deviation was calculated for all 4 of the first and then second readings for all 12 cases, providing 24 average deviations to enable calculation of the mean measurement difference and standard error in millimetres. The student *t*-test was used to test the hypothesis that there would be a difference in the variation for smaller tumours, less than 15 mm in diameter, compared to larger tumours, 15 mm or more in diameter. Intraclass correlation was employed to determine the reliability of the measurements, where it is assumed that a correlation value of more than 0.7 is reliable. Intraclass correlation and 95 per cent confidence intervals (CIs) were calculated using an array of all the readings obtained for each of the measured structures.

Statistics were calculated using Microsoft Excel® for Mac 2011 software, version 14.5.4, with the Real Statistics Resource Pack for Excel 2011 add-in.

Results

Table I presents mean measurement deviation with standard error for all 12 cases. Inter-observer difference for tumour diameter averaged 0.33 ± 0.04 mm (range, 0.0–0.8 mm), and intra-observer difference averaged 0.17 ± 0.03 mm (range, 0.0–0.8 mm). Figure 1 demonstrates the variation in all the readings of maximal tumour diameter as a scatter plot. A student's *t*-test investigating the difference in mean differences between inter- and intra-observer measurements demonstrated more consistency between readings by the same radiologist, with average lower differences of 0.17 mm (95 per cent CI = 0.27–0.06, $p = 0.002$).

The hypothesis that a significant difference in inter-observer variability existed for smaller tumours versus

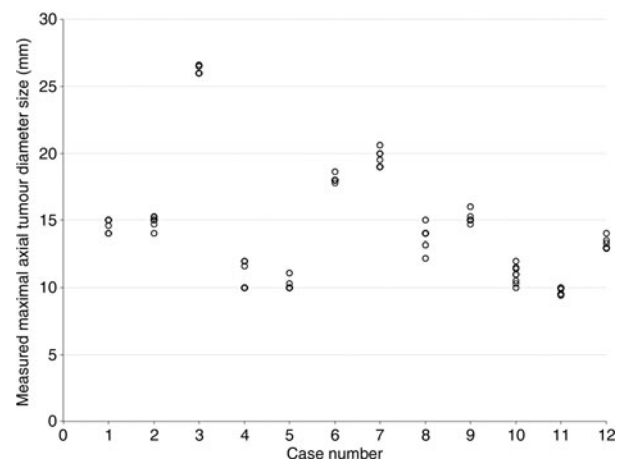


FIG. 1

Measurement variation in 12 cases of cerebellopontine angle tumours.

larger tumours was tested with a student *t*-test. Despite there being a trend with a larger mean difference for the smaller tumours (less than 15 mm) at 0.44 ± 0.07 mm (range, 0.11–0.86 mm), compared to the larger tumours (15 mm or more) at 0.30 ± 0.05 mm (range, 0.08–0.65 mm), the results were not significantly different ($p = 0.09$).

Test results for reliability using intraclass correlation are presented in Table II. Inter-observer reliability for maximal tumour diameter was 0.99 (95 per cent CI = 0.97–0.99). These results indicate that the measurements between radiologists were extremely reliable. Similarly, intra-observer reliability, at 0.99 (95 per cent CI = 0.99–1.00), were extremely reliable.

Discussion

According to a recent study, intervention is generally not undertaken until the cerebellopontine angle moiety of the tumour reaches 15 mm.⁶ If we take a hypothetical case of a tumour that grows from 4 mm to 6 mm in diameter and one considers the spherical volume of a sphere ($\frac{4}{3} \pi r^3$), such a change represents a volume increase of more than three times. Some studies have found that there is an existing error of 1–2 mm in inter-observer tumour measurement.^{1–3}

TABLE I
MEAN MEASUREMENT DEVIATION IN CEREBELLOPONTINE ANGLE TUMOUR MRI ASSESSMENT

Measurement	Mean measurement deviation (mm)	
	Mean ± SE	Range
Inter-observer		
– Tumour diameter	0.33 ± 0.04	0.0–0.8
– Ipsilateral peduncle	0.40 ± 0.04	0.04–0.93
– Contralateral peduncle	0.38 ± 0.03	0.11–0.75
Intra-observer		
– Tumour diameter	0.17 ± 0.03	0.0–0.8
– Ipsilateral peduncle	0.17 ± 0.04	0.0–1.0
– Contralateral peduncle	0.14 ± 0.03	0.0–1.0

MRI = magnetic resonance imaging; SE = standard error

TABLE II
INTRACLASS CORRELATION OF CEREBELLOPONTINE ANGLE TUMOUR MRI ASSESSMENT

Measurement	Intraclass correlation	95% CI
Inter-observer		
– Tumour diameter	0.99	0.97–0.99
– Ipsilateral peduncle	0.94	0.89–0.97
– Contralateral peduncle	0.90	0.82–0.95
Intra-observer		
– Tumour diameter	0.99	0.99–1.00
– Ipsilateral peduncle	0.97	0.95–0.98
– Contralateral peduncle	0.96	0.92–0.98

MRI = magnetic resonance imaging; CI = confidence interval

Of course, no major clinical decision would be made on such a volume increase in a tumour of this size, but there are significant issues of patient concern and anxiety.

This study indicates very little test–retest change and very little inter-observer difference. Such results, unexpected by the senior author, have not been shown in other studies.^{2–4} In order to avoid surgery that proves unnecessary, and to similarly avoid alarming patients about apparent tumour growth, the authors believe that if an expectant policy is to be undertaken, serial imaging should be carried out in the same institution by experienced neuroradiologists.

While the results of this study are reassuring, it is what is to be expected from highly trained neuroradiologists who use the same machine and who work very closely with the clinicians involved. The clinicians themselves also need to be confident about their ability to interpret the cases placed before them.

Clinicians should give precise instructions to their radiologists requesting that the cerebellopontine angle moiety be individually measured at the point of maximum horizontal extent.⁷

The slice thickness of the MRI scans used in this study was 0.5 mm; to our knowledge, this is the smallest interval that has been used to test for inter-observer variability. The next closest study had used a 1.0 mm slice thickness, yielding similar mean readings but a comparatively larger error.²

- **Errors of 1–2 mm on two-dimensional (2D) measurements have been demonstrated**
- **Volumetric measurements are superior to 2D measurements in determining growth**
- **However, volumetric measurements are time-consuming, dependent on magnetic resonance imaging sequences and are still susceptible to human error**
- **Despite the errors, 2D measurements are reliable and practical for clinical use**
- **There is innate bias associated with intra-observer readings and different measuring protocols**
- **The same centre and radiologist should ideally monitor tumour progression to enhance decision making**

While there are claims that volumetric analysis is important,⁵ this is an expensive technique that is not available to everybody, and is in itself susceptible to human error.^{4,8} These authors believe that the assessment of middle cerebellar peduncle compression (as conducted in this study) is probably the most important factor leading to intervention. While unproven, we propose that slice thickness influences the accuracy of axial measurements.

Acknowledgements

The authors would like to acknowledge the following radiologists who participated in our study: John Ly, Pascal Bou-Haidar, Sebastian Fung and Yael Barnett.

References

- 1 van de Langenberg R, de Bondt B, Nelemans P, Baumert B, Stokroos R. Follow-up assessment of vestibular schwannomas: volume quantification versus two-dimensional measurements. *Neuroradiology* 2009;**51**:517–24
- 2 Hougaard D, Norgaard A, Pedersen T, Bibby B, Ovesen T. Is a redefinition of the growth criteria of vestibular schwannomas needed? *Am J Otolaryngol* 2014;**35**:192–7
- 3 Marshall AH, Owen VM, Nikolopoulos TP, O'Donoghue GM. Acoustic schwannomas: awareness of radiologic error will reduce unnecessary treatment. *Otol Neurotol* 2005;**26**:512–15
- 4 Cross J, Baguley D, Antoun N, Moffat D, Prevost A. Reproducibility of volume measurements of vestibular schwannomas - a preliminary study. *Clin Otolaryngol* 2006;**31**:123–9
- 5 Harris GJ, Plotkin SR, Maccollin M, Bhat S, Urban T, Lev MH *et al.* Three-dimensional volumetrics for tracking vestibular schwannoma growth in neurofibromatosis type II. *Neurosurgery* 2008;**62**:1314–20
- 6 Jufas N, Flanagan S, Biggs N, Chang P, Fagan P. Quality of life in vestibular schwannoma patients managed by surgical or conservative approaches. *Otol Neurotol* 2015;**36**:1245–54
- 7 Tanaka Y, Hongo K, Tada T, Kobayashi S. What is the best method for reporting tumor diameter in vestibular schwannoma? *Neurosurgery* 2003;**53**:634–8
- 8 Yamada I, Tsunoda A, Noguchi Y, Komatsuzaki A, Shibuya H. Tumor volume measurements of acoustic neuromas with three-dimensional constructive interference in steady state and conventional spin-echo MR imaging. *J Magn Reson Imaging* 2000;**12**: 826–32

Address for correspondence:

Sheng Rong Teh,
Room 1.16,
215A Anzac Parade,
Kensington,
NSW 2033, Australia

E-mail: srong_teh@hotmail.com

Prof P Fagan takes responsibility for the integrity of the content of the paper

Competing interests: None declared