The Journal of International Advanced Otology

J Int Adv Otol 2025; 21: 1-9 • DOI: 10.5152/iao.2025.241693

Original Article

# The Real-World Impact of Vestibular Schwannoma Fully Automated Volume Measures on the Evaluation of Size Change and Clinical Management Outcomes in a Multidisciplinary Meeting Setting

**Steve Connor[1], Navodini Wijethilake[1], Anna Oviedova[2], Rebecca Burger[2], Marina Ivory[1], Tom Vercauteren[1], Jonathan Shapey[1]**

[1]King's College London School of Biomedical Engineering and Imaging Sciences, London, United Kingdom
[2]Department of Neurosurgery, King's College Hospital, London, United Kingdom

ORCID iDs of the authors: S.C. 0000-0001-5502-4972, N.W. 0000-0001-9620-8233, A.O. 0000-0001-8404-3961, R.B. 0000-0002-3591-9178, M.I. 0000-0001-5583-6700, T.V. 0000-0003-1794-0456, J.S. 0000-0003-0291-348X.

Cite this article as: Connor S, Wijethilake N, Oviedova A, et al. The real-world impact of vestibular schwannoma fully automated volume measures on the evaluation of size change and clinical management outcomes in a multidisciplinary meeting setting. *J Int Adv Otol.* 2025, 21(2), 1693, doi: 10.5152/iao.2025.241693.

**BACKGROUND:** Vestibular schwannoma (VS) management decisions are made within multidisciplinary meetings (MDMs). The improved accuracy of volumetric compared to linear tumor measurements is well-recognized, but current volumetric evaluation methods are too time-intensive. The aim was to determine if the availability of fully automated volumetric tumor measures during MDM preparation resulted in different radiological outcomes compared to a standard approach with linear dimensions, and whether this impacted the clinical management decisions.

**METHODS:** A prospective cohort study evaluated 50 adult patients (mean age 64.6, SD 12.8; 24 male, 26 female) with unilateral sporadic VS. Two simulated MDMs were convened using different methods to measure tumor size during radiology preparation: MDM-mlm used linear tumor dimensions, while MDM-avm was provided with fully automated deep learning-based volume measurements. Interval changes in VS size from the index to final and penultimate to final magnetic resonance imaging (MRI) studies defined the radiological outcomes. The subsequent clinical MDM outcomes were classified. Wilcoxon signed rank tests compared the radiological classification of VS size change and the management outcomes between the MDM-mlm and the MDM-avm.

**RESULTS:** The 57 interval MRI comparisons in 33 patients showed a significant difference in the classification of VS size change between the MDM-mlm and MDM-avm for all intervals ($z = 2.49$, $P = .01$). However, there was no significant difference in the resulting management decisions between the 2 MDMs ($z = 0.30$, $P = .76$).

**CONCLUSION:** Provision of fully automated VS volume measurements to "real-world" MDM preparation significantly impacted the radiological classification of VS size change but did not influence management decisions.

**KEYWORDS:** case report, intracranial schwannomatosis, optic nerve, schwannoma, supraorbital approach

## INTRODUCTION

Vestibular schwannomas (VS) are benign tumors of the eighth cranial nerve sheath. The lifetime risk of being diagnosed with a VS likely exceeds 1 in 500.[1] Since most tumors are indolent and are increasingly detected when small, an initial conservative management strategy with observation and magnetic resonance imaging (MRI) monitoring is often proposed. The intensity of serial imaging and the interval between MRI studies will depend on tumor size and location, stability, and duration of follow up.[2] Determining that such a "watch and wait" approach has failed and that treatment with microsurgery or stereotactic radiotherapy is required is multifactorial, and will be influenced by patient age, clinical features, and co-morbidities. However, demonstrating tumor growth on sequential MRI is a key factor in deciding whether therapeutic intervention is appropriate.

**Corresponding author**: Steve Connor, e-mail: steve.connor@kcl.ac.uk
**Received:** September 5, 2024 • **Revision Requested:** October 26, 2024 • **Last Revision Received:** November 8, 2024 • **Accepted:** December 4, 2024 • **Publication Date:** March 24, 2025
Available online at www.advancedotology.org

Monitoring of VS growth has traditionally been performed with 2-dimensional linear measurements.[3,4] However, volumetric evaluation is more reliable and sensitive for the detection of VS growth.[5-13] Tumor volumes may be measured with manual segmentation, but this is labor-intensive and subjective,[7,8,11] so precluding its routine clinical application. An alternative approach is with semi-automated volumetric segmentation tools, which allow for more rapid and repeatable evaluation, but again there remains a requirement for operator interactions and alterations of the segmentation.[14] Fully automated image segmentation aims to smooth workflow by contouring the tumor volume without the need for further manual adjustments. Such fully automated approaches are now possible with deep learning (DL) techniques, and these have been applied to the segmentation of VS across a range of MRI sequences and clinical datasets.[15-20]

Information on VS size and growth is of vital importance to multidisciplinary team meetings (MDMs) or tumor boards in order to allow management decisions. However, MDMs and radiology departments are struggling to provide sufficient time for radiologists to prepare for MDMs and to perform tumor measurements as the case load increases.[1,21] The provision of fully automated measurements may usefully augment the evaluation of serial VS size changes by the radiologist and reduce errors resulting from increased time pressures.[22] The implementation of DL-generated automated measurements of VS size within the MDM setting has been described;[23] but the clinical impact of applying fully automated volumetric measurements remains to be explored. As such volumetric tumor analysis becomes more feasible to apply in clinical practice, it is important that the effect on radiological and management outcomes in such real-world settings is evaluated.

Our primary objective was to determine whether the availability of fully automated volume measures resulted in different classifications of VS size changes as compared to manual linear dimensions, and whether this impacted clinical management decisions in a real-world MDM setting. Our secondary objectives were to evaluate whether manual linear dimensions performed outside the MDM setting resulted in a different classification of VS size change compared to those performed within the time-pressured environment of MDM preparation, and to assess whether manual linear measures were able to predict changes in the fully automated volume measures. A STARD checklist guided this work.

## MAIN POINTS

- The impact of deep-learning-based fully automated volume measurements on radiological outcomes and clinical management was evaluated in a real-world multidisciplinary meeting (MDM) setting.
- There was a difference in the classification of interval vestibular schwannoma size changes between an MDM applying linear tumor dimensions, and an MDM where fully automated deep learning-based volume measurements were provided.
- The differences in radiological outcomes did not translate into a difference in the MDM clinical management decisions, although this may be a result of the small patient cohort.
- The deep-learning-based model may usefully augment the radiologist interpretive skills while improving workflow and adding certainty; however, its real benefits will only be achieved when there is a clear impact on patient care.

## METHODS

### Patients

The study was approved by the Research Ethics Committee (Approval:22/NS/0160(AIMBraTS); date June 12, 2022), and the requirement for patient consent was waived. This prospective cohort study evaluated 50 adult patients with unilateral sporadic vestibular schwannoma (Figure 1). Patients were curated from a list of 200 consecutive patients referred to a tertiary neuroscience center MDM between December 2009 and September 2012 and were chosen to be representative of our unit's typical MDM composition in terms of the number, surveillance period, and rate of intervention. The final MRI studies were performed between 9/1/12 and 15/3/21. The mean interval between the initial and final MRI study was 54.7 (standard deviation (SD) 31.9) months in the 50 patients, and a total of 187 MRI studies were performed. This included 4 post-operative and 11 post-stereotactic radiosurgery (SRS) tumors, with index tumors being intrameatal (IM) in 13 patients, extrameatal (EM) in 16 patients, and post-operative remnants in 4 patients (Table 1).

### Simulated Multidisciplinary Meetings

An MDM using manual linear measurements (MDM-mlm) and an MDM with the availability of automatic volumetric assessments (MDM-avm) were convened 35 days apart. The same cases were considered in both MDMs, but the order of cases was randomized. In line with standard practice, both allowed for up to 2.5 hours of preparation of the cases by the neuroradiologist (SC with 25 years of subspecialty experience) prior to the MDM. The time taken to evaluate each case during the preparation for each MDM was documented.

### MDM-mlm Preparation

The preparation for the MDM-mlm incorporated the local "standard of care." All sequences were reviewed at each time point, and the highest resolution MRI sequences, T2w or gadolinium enhanced (Gd) T1w MRI sequences, were selected to acquire measurements. Digital calipers were placed using the Picture Archiving & Communications System (PACS) workstation (Sectra workstation, Sectra AB, Sweden) on a section of the tumor that was perceived to demonstrate the maximal linear dimension. The linear measurements were always performed in the axial plane; however, subsequent images were reformatted where possible to correct for differences in obliquity between interval MRI studies. The maximum whole tumor axial dimension was obtained when the index tumor was purely IM, and the maximum EM axial dimension (in any direction) was measured when the index tumor EM component was larger than the porus acusticus (Figure 2). The radiological outcomes were recorded for interval changes in VS linear dimensions from the index to the final (I-F) MRI and from the penultimate (the imaging study prior to the final study) to the final (P-F) MRI studies based on measurements performed to the nearest 0.1 mm. If VS surgery was performed after the initial MRI study, then the first post-operative MRI was designated as the index MRI. The radiological outcome was defined as either definite (>2 mm) increase,[4] equivocal (1-2 mm) increase, stable (<1 mm change), equivocal (1-2 mm) decrease, or definite (>2 mm) decrease. The definitions for definite change in linear measurements were based on consensus guidelines, with the measurement thresholds halved

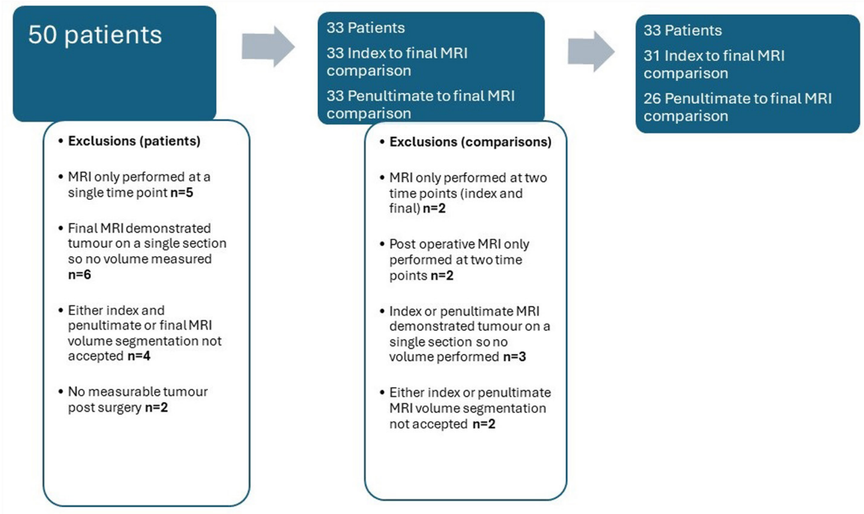


**Figure 1.** Flow chart demonstrating selection process and exclusions for final cohort.

to define equivocal changes.[4] In accordance with routine practice, the VS linear measurement (to the nearest 1 mm), radiological outcome, and free text for any additional relevant imaging features were recorded on a structured form to present at the MDM-mlm.

### MDM-avm Preparation

The preparation for the second MDM-avm was augmented by a report providing automated volume measurements for each VS MRI derived from DL-based segmentations, using in-house software trained on a separate dataset.[18,23,24] Segmentation was performed on the highest resolution Gd T1w MRI sequence or the high-resolution T2w imaging when Gd T1w imaging was not available. Automatically generated volume segmentations were displayed as 2D masks on serial sections, and interval changes were depicted in a bar plot. The neuroradiologist evaluated the segmentation alongside the images on PACS for index, penultimate, and final VS volumes and decided whether each was acceptable (Figure 3). Interval changes in VS % volume (% Δ vol) and absolute volume for I-F and P-F defined the radiological outcome.[4] These were recorded as definite (>20% or 1.2 $cm^3$) increase, equivocal (10%-20% or 0.6-1.2 $cm^3$) increase, stable (<10% or 0.6 $cm^3$ change), equivocal (10%-20% or 0.6-1.2 $cm^3$) decrease, or definite (>20% or 1.2 $cm^3$) decrease. The definitions for definite change in volume measurements were based on consensus guidelines, with the measurement thresholds again halved to define equivocal changes.[4]

### Multidisciplinary Meetings and Clinical Management Outcomes

Multidisciplinary meetings immediately followed the preparation and were conducted in an online and in-person hybrid format as per normal practice, with 9 clinical staff (2 skull base neurosurgeons, 2

**Table 1.** Patient Demographics and Tumor Characteristics for All MDM Cases and the Final Study Cohort

| | | All MDM Cases | Study Cohort |
|---|---|---|---|
| Number of patients | | 50 | 33 |
| Number of comparisons | Index-final | 47 | 31 |
| | Penultimate-final | 45 | 26 |
| Mean (SD) interval between MRI scans | Index-final | 54.7 (31.9) | 52.3 (32.5) |
| | Penultimate-final | 24.9 (19.5) | 23.9 (22.6) |
| Mean age (SD) | | 64.6 (12.8) | 65.1 (13.7) |
| Sex (M/F) | | 24/26 | 14/19 |
| Treatment interventions before final MRI | | Surgery 8 (16%)<br>12 SRS (24%) | Surgery 4 (12%)<br>SRS 11 (33%) |
| Number of index tumors at each location | Intrameatal | 23 | 13 |
| | Extrameatal | 23 | 16 |
| | Index MRI of post-operative remnant | 4 | 4 |
| Mean linear dimension (SD) at each location | Intrameatal | 9.0 cm (3.4 cm)<br>whole tumor | 10.8 cm (4.1 cm)<br>whole tumor |
| | Extrameatal | 18.4 cm (9.1cm)<br>extrameatal tumor | 14.3 cm (6.1 cm)<br>extrameatal tumor |
| | Index MRI of post-operative remnant | 17.8 cm (9.2 cm) | 17.8 cm (9.2 cm) |

SRS, stereotactic radiosurgery, MDM, multidisciplinary meeting.

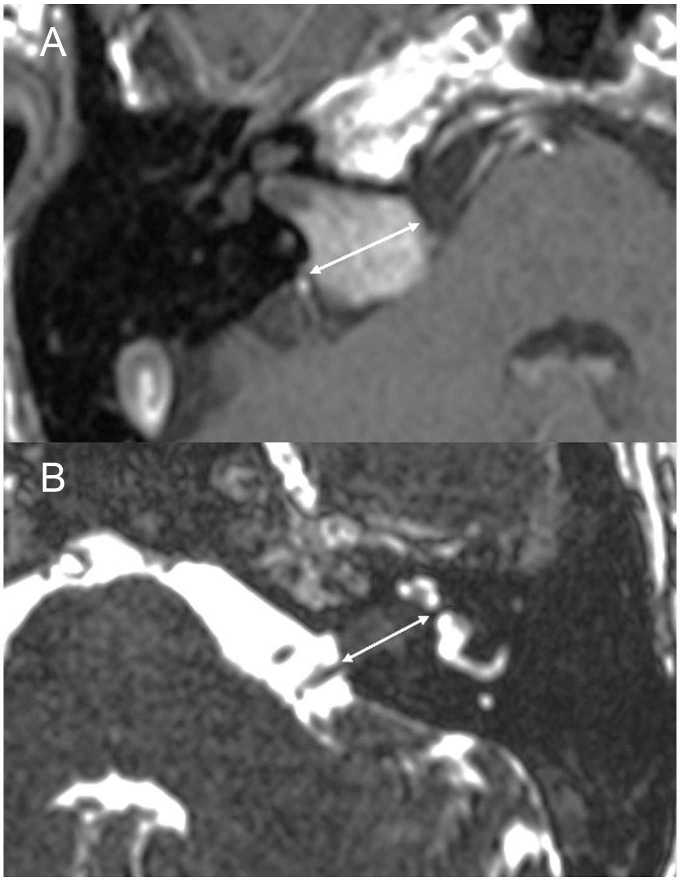


**Figure 2.** Linear dimensions evaluated for intrameatal and extrameatal tumors. (A) Post gadolinium T1w image with 3 mm axial sections demonstrates an extrameatal component larger than the porus acusticus, so a linear measurement is performed of the maximum axial dimension of the extrameatal component (in any direction). (B) Three-dimensional T2w image with 0.7 mm slice thickness demonstrates a purely intrameatal component, so the linear measurement is performed of the whole tumor.

clinical oncologists, 2 clinical nurse specialists, and 3 neurosurgical fellows). Each case was demonstrated on the PACS system through screen sharing with a discussion of the recorded tumor dimensions and interval changes. During the MDM-avm, members also referred to the summary automated report to assist their decision-making with an emphasis on % Δ vol. The proposed management outcome was classified as: 1) discharge from follow-up, 2) surveillance (increased interval for next MRI), 3) intensive surveillance (decreased interval for next MRI), or 4) intervention (surgery or stereotactic radiosurgery).

### Final Study Cohort and Analysis of Linear Dimensions Outside the Multidisciplinary Meeting Setting

Interval MRI comparisons were excluded when 1) VS volume segmentation was not available (due to limited time points), 2) VS volume could not be calculated (the tumor was depicted on a single axial section), or 3) VS volume was unacceptable (radiologist judged it did not to correspond to the contour of the tumor).

Eleven months after the initial MDM-mlm, further manual linear dimensions were performed for this final study cohort outside the MDM setting. These were performed to the nearest 0.1 mm, and radiological outcomes were classified at the same MRI time points and on the same sequences, while blinded to initial measurements.

### Statistical Analysis

Statistical analyses were performed using SPSS® Statistics 27.0 (IBM SPSS Corp.; Armonk, NY, USA).

A *P* value <.05 was considered statistically significant. Normality of data was evaluated with the Shapiro–Wilk test, with non-parametric tests performed when $P < .05$. Descriptive statistics documented the mean and SD for normally distributed data and recorded the median and interquartile range (IQR) for data that was not normally distributed.

Wilcoxon signed rank tests and Kendall's Tau compared the 3 classifications of VS size change (stable, equivocal, or definite), and the 5 management outcomes (discharge from follow-up, surveillance, intensive surveillance, and intervention) between those documented at the MDM-mlm and those recorded at the MDM-avm.

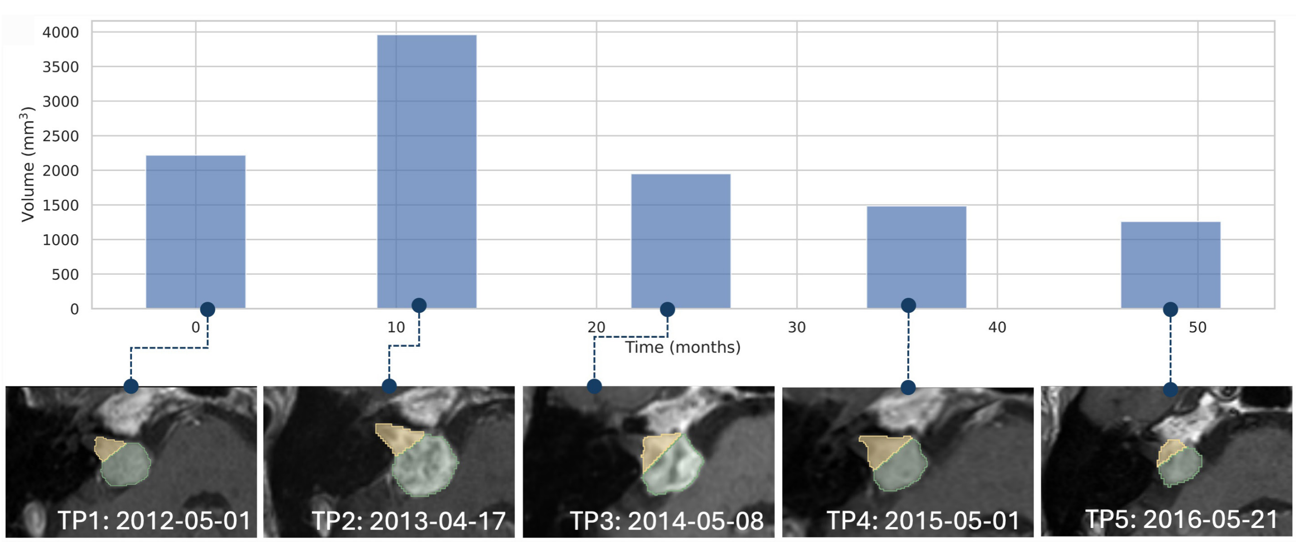


**Figure 3.** Example of volume segmentations and report available for MDM-avm preparation. Five serial gadolinium-enhanced T1w MRI studies demonstrate an increase in VS volume and subsequent regression following stereotactic radiotherapy. The index MRI at TP (time point) 1 is the same as that depicting the linear measurement in Figure 2A. The segmentations were considered acceptable, and volume comparisons were performed from both TP1 to TP5 and TP4 to TP5.

Weighted Cohen's kappa evaluated the reliability of the radiological outcomes (definite increase, equivocal increase, stable, equivocal decrease, definite decrease) recorded from linear measures within and outside the MDM setting. The Wilcoxon signed-rank test compared the classification of VS size change derived from linear dimension measures outside the MDM setting with those during the MDM-mlm preparation and with those obtained from MDM-avm volume measures.

Spearman rank correlation coefficient established the relationship between changes in the linear interval changes in the manual linear dimensions and the volumetric change in VS size. Volume change is proportional to the cube of linear diameter change, so to evaluate linear growth in a comparable manner to volumetric growth, % linear interval changes were cubed (% Δ $\text{linear}^3$) for comparison with the automated % volumetric changes (% Δ vol). The sensitivity and specificity of equivocal (1 mm) and definite (2 mm) increases in linear dimensions for the detection of definite volumetric tumor growth were also evaluated.

## RESULTS

### Descriptive Data for Cohort and Multidisciplinary Meetings

Of the 50 adult patients included in the MDM cohort (mean age 64.6 SD 12.8; 24 male, 26 female), there were 17 patients in whom both the I-F MRI and P-F MRI comparison were excluded and 9 patients in whom one comparison was excluded. Therefore, the study cohort included 33 patients (mean age 65.1, SD 13.7; 16 males, 19 females) with 57 interval MRI comparisons (Figure 1). The median index tumor volume [IQR] was 610.3 [253.6-2374.0] $\text{mm}^3$.

The linear measurement comparisons in MDM-mlm and outside the MDM setting were performed on serial MRI sequences with weighting and section thickness as documented in Table 2.

The time taken to review imaging for each of the 33 patients was a mean 2 minutes 50 seconds (SD 1 minute 4 seconds) for the MDM-mlm preparation and 3 minutes 3 seconds (SD 1 minute 31 seconds) for the MDM-avm preparation. There was no significant difference between the MDM-mlm and the MDM-avm for the time required for the preparation of each case ($z=1.36$, $P=.10$). Additional recorded findings not reflected by the MDM-avm volumetric assessment included significant interval internal necrosis (n = 3) and the presence of non-enhancing peritumoral cysts (n = 1) in tumors with surveillance as the management strategy.

**Table 2.** MRI Sequences and Section Thickness for the Linear Dimensions (n = 57)

| Index or Penultimate MRI Slice Thickness (Sequence) | Final MRI Slice Thickness (Sequence) | Number |
|---|---|---|
| 3 mm (Gd T1w) | <1 mm (Gd T1w) | 20 |
| 3 mm (Gd T1w) | 2-3 mm (Gd T1w) | 14 |
| 0.7 mm (T2w) | 0.7 mm (T2w) | 10 |
| <1 mm (Gd T1w) | <1 mm (Gd T1w) | 5 |
| 0.7 mm (T2w) | <1 mm (Gd T1w) | 4 |
| 0.7 mm (T2w) | 2-3 mm (Gd T1w) | 4 |

Gd, post-gadolinium.

### Multidisciplinary Meeting-mlm and Multidisciplinary Meeting-avm Classifications of Vestibular Schwannomas Size Change and Management Outcomes

There was a significant difference in the classification of VS size change recorded in the MDM-mlm compared to the MDM-avm, both for I-F alone ($z=2.44$, $P=.02$) and all intervals ($z=2.49$, $P=.01$). There was also no significant correlation between the classification of VS size change obtained from MDM-mlm linear dimensions and MDM-avm volumetric measures for these interval comparisons ($\tau=0.502$, $P=.003$; $\tau=0.403$, $P<.001$). There were 24 interval comparisons judged to show definite changes in the MDM-mlm while 35 interval comparisons with definite changes in the MDM-avm. There was no difference between the management decisions resulting from the 2 MDMs and they were significantly correlated ($z=0.30$, $P=.76$; $\tau=0.716$, $P=.357$) (Figure 4, Table 3).

### Radiological Outcomes Obtained Outside the Multidisciplinary Meeting Setting

Weighted Cohen's kappa ($\kappa$) demonstrated very good agreement between the 5 radiological outcomes documented during MDM-mlm preparation and those recorded outside the MDM setting for all interval comparisons (weighted $\kappa=0.848$; 95% CI, 0.748-0.949).

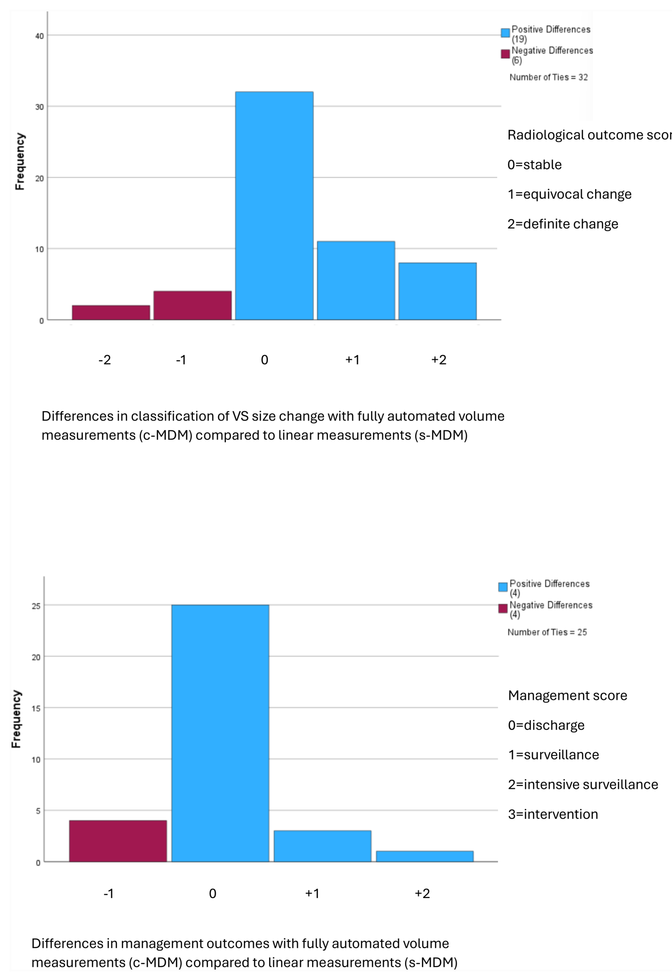


**Figure 4.** Chart demonstrating differences in VS size change classification and management outcome between MDM-mlm and MDM-avm.

**Table 3.** Comparison of Radiological Management Outcomes Between the MDM-mlm and MDM-avm

| Comparison | Classification of VS Size Change | Number | MDM-mlm (Linear) | MDM-avm (Volume) | Wilcoxon Signed Rank | Kendall's Tau |
|---|---|---|---|---|---|---|
| **Radiological outcome** | | | | | | |
| Index to final MRI | Stable | 31 intervals | 8 | 3 | $P=.02$ $z=2.44$ | $P=.003$ $\tau=0.502$ |
| | Equivocal | | 2 increase<br>4 decrease | 3 increase<br>1 decrease | | |
| | Definite | | 10 increase<br>7 decrease | 13 increase<br>11 decrease | | |
| Penultimate to final MRI | Stable | 26 intervals | 12 | 9 | $P=.24$ $z=1.19$ | $P=.249$ $\tau=0.211$ |
| | Equivocal | | 3 increase<br>4 decrease | 3 increase<br>3 decrease | | |
| | Definite | | 3 increase<br>4 decrease | 5 increase<br>6 decrease | | |
| All intervals | Stable | 57 intervals | 20 | 12 | $P=.01$ $z=2.49$ | $P<.001$ $\tau=0.403$ |
| | Equivocal | | 5 increase<br>8 decrease | 6 increase<br>4 decrease | | |
| | Definite | | 13 increase<br>11 decrease | 18 increase<br>17 decrease | | |
| Management outcome | Discharge | 33 patients | 3 | 4 | $P=.763$ $z=0.30$ | $P=.357$ $\tau=0.716$ |
| | Surveillance | | 17 | 15 | | |
| | Intensive surveillance | | 9 | 9 | | |
| | Intervention | | 4 | 5 | | |

There was no difference between the 3 classifications of VS size change based on linear dimensions acquired during MDM-mlm preparation and those performed outside the MDM setting ($z=133.0$, $P=.088$). The classification of VS size change using linear dimensions outside the MDM setting remained significantly different from that obtained when applying the MDM-avm volume measures ($z=64.0$, $P=.001$).

### Relationships Between Linear Dimension and Fully Automated Volume Changes

Since the radiological outcome was reflected by interval changes in VS % volume (% Δ vol) in all cases and absolute volume changes in only 48/57 cases, the % Δ vol was used for a comparison with the linear measures. There was a significant positive relationship between changes in the cube of the manual linear dimensions (% Δ linear$^3$) (mean 15.46 SD 85.62%) and the % Δ vol (median 9.90 IQR [−90.4 to 110.2] %) for all intervals ($r_s=0.76$, $P<.001$) (Figure 5). A threshold of 1 mm (equivocal) increase in linear dimension predicted definite (>20%) % Δ vol with sensitivity 83.3%, specificity 71.8%, and diagnostic odds ratio (DOR) of 12.73 (95% CI 3.07-52.78), while a threshold of 2 mm (definite) increase predicted volumetric tumor growth with sensitivity 67.7%, specificity 87.2%, and DOR of 13.60 (95% CI 3.50-52.83).

## DISCUSSION

This study of real-world application of fully automated VS volume changes in the MDM setting demonstrated a significant difference in the classification of all interval VS size changes when compared with standard linear measurements ($P=.01$) and a greater number of cases were classified as definite change, although this did not significantly impact management decisions ($P=.76$). The evaluation of linear dimension changes outside the MDM setting did demonstrate very good agreement with those recorded during MDM preparation ($\kappa=0.848$), and there remained a significant difference in the classification of volume changes in VS size ($P=.001$). The % Δ linear$^3$ significantly correlated with the % Δ vol ($P<.001$). However, the current definition of definite linear dimension change (2 mm)[4] was shown to be insensitive (67.7%) to definite volume change (>20%). There was a similar ability of 1 mm (DOR 12.73) and 2 mm (DOR 13.60) thresholds for linear dimension change to discriminate definite % change in VS volume.

Artificial intelligence (AI) has been applied to the streamlining of MDMs and has the potential to aid clinical decision making.[25-27] Radiologists preparing for MDMs are confronted by growing case numbers and serial imaging,[21] and while our study did not show any significant difference in the radiological outcomes obtained within

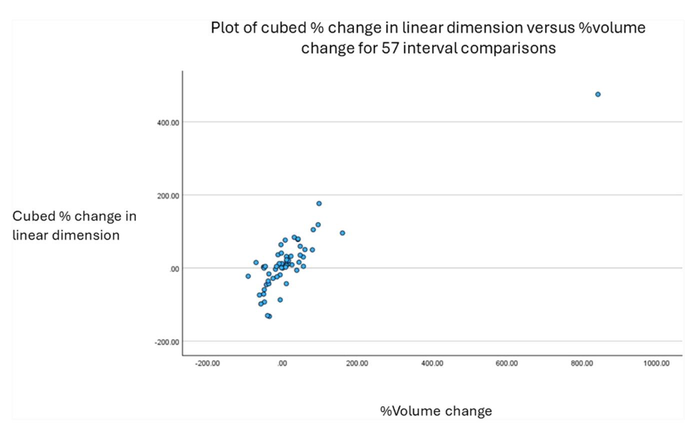


**Figure 5.** Plot of cubed % change in linear dimension versus % volume change for 57 interval comparisons.

and outside the MDM settings, the time pressure during the MDM preparation has the potential to result in diagnostic errors and variable interpretation. The review of sequential MRI studies to determine fluctuations in VS size is a key component of the skull base MDM, and it is recognized that volumetric changes provide the most reliable and accurate assessment. In this context, radiologists may benefit from AI applications such as DL-based tumor segmentation to aid their interpretative skills.[27] Most studies of DL models for VS volume segmentation have focused on internal cohorts, although some have undergone external validation from heterogeneous multi-institution data.[15-17,19,20] There has been limited application of DL-based volumetric segmentation of tumors in other real-world settings.[28] Testing within the MDM environment could aid the acceptance and penetration of these tools into clinical practice and overcome skepticism concerning their implementation.

The complementary nature of radiologist and AI input into MDM preparation was demonstrated by the requirement to reject unrepresentative volume segmentations, which precluded at least one of the I-F or P-F volumetric comparisons in 7 patients. It is also notable that there were 4 cases in which significant tumor-related features required communication by the radiologist. While the process of verifying the volume segmentations resulted in an overall increase in MDM preparation time for the study cohort, no prior training was obtained, and it is anticipated that this would decrease with familiarity with the process.

Although there was a clear impact of DL-based volumetric tumor changes on radiological outcomes, there was no significant impact on the clinical management strategy.[23] This implies that clear size changes recognizable by both linear and volume measures are currently required to prompt intervention. It should be appreciated that therapeutic choices are multifactorial and also depend on the rate of growth, hearing symptoms, demographics, and comorbidities. While the same patients were reviewed in each MDM and hence "patient factors" were consistent, it is possible that they were not considered with equal weighting, which may have differentially influenced the management decisions. It should also be noted that our elderly cohort (mean age 65.1) with small tumor volumes (median 610.3 $mm^3$) could result in a bias towards surveillance rather than intervention, and it is possible that there would be a greater impact on clinical management if a younger population with larger tumors had been studied.

Despite our study findings, it is expected that automated volumetric evaluation will become embedded within clinical practice through its gains in precision and efficiency. It is expected that the evolution of robust, generalized DL algorithms and their application to contemporary high-resolution volumetric MRI sequences will improve the accuracy, reliability, and acceptability of segmentations. To have a greater impact on the development of individualized MRI surveillance protocols, there will also need to be greater consensus on the definition of volumetric tumor growth, and this will be enabled by accrual of volumetric data. Accurate volumetric analysis will be of particular benefit in the setting of neurofibromatosis type 2 tumors, which have complex growth patterns that are poorly defined by linear dimensions. Pre-treatment VS volume changes, post-treatment residual volume, and patient factors (e.g., age) may contribute to multivariable predictive models to determine the propensity to tumor re-growth.[29] In addition to the evaluation of tumor volume, the ability to perform automated segmentation on MRI may aid prognostication and assessment of treatment response through the interrogation of radiomics features,[30] and physiological properties such as perfusion,[31] while it may contribute to treatment planning through tumor contouring for radiotherapy targets.[32] The incorporation of such DL-based automation within MDM practice presents challenges that are common to the adoption of any artificial intelligence technologies. In particular, there is a need to overcome mistrust of the DL algorithm's reliability among MDM members, perceived threat to the agency, autonomy, and expertise of the radiologist, and concerns about the managerial justification for introducing these automated tools.[33]

Previous studies have also shown decreased sensitivity using linear dimension criteria compared to a >20% volume definition of VS growth,[6,8,10,13] potentially resulting in a bias to non-intervention. Morris et al[8] demonstrated a sensitivity of 86% for the discrimination of volumetric progression when using a >2 mm linear dimension criterion in 61 NF2-associated VS. However, they applied a range of standardized linear measurements rather than a single axial dimension. Other studies have demonstrated sensitivities of 71%,[6] 56%,[9] and 57%[13] when applying a >20% linear dimension increase to the prediction of a >20% volume VS progression. Harris et al[6] and Walz et al[13] also found that the % Δ linear[3] underestimated % Δ vol by 50% and 73%, respectively. Our contrasting observation of % Δ linear[3] exceeding % Δ vol may relate to the inclusion of 13/33 IM tumors in which the volume increase is limited by the confines of the IAC.

The study has limitations that should be considered. Firstly, there were constraints in data collection during the "standard of care" MDM-mlm preparation. For instance, the MDM-mlm documented the radiological outcome based on linear measurements obtained in 0.1 mm increments, but only the radiological outcome and the tumor dimensions (to the nearest 1 mm) were recorded. Thus, the evaluation of intra-observer reliability within and outside the MDM setting was confined to variations in radiological outcomes rather than absolute linear measures. Differences in linear dimensions have previously shown poor reliability relative to volumetric changes.[5,7,10,11] It also did not allow for multiple linear measurements in different planes, which better predict volumetric tumor growth.[9] Secondly, there was potential for bias by the high rate of exclusions, although the final study cohort was shown to be broadly representative of the initial patient group and reflected a standard MDM case mix. Thirdly, the patient sample was accrued from those referred up to 15 years previously, resulting in exclusions due to a number of thick-section index studies precluding volume segmentation. Finally, the robustness of the DL-based volume changes should be addressed in the context of heterogeneous imaging sequences, thick-section imaging, small IM tumors, and post-operative remnants.[34] While the DL segmentation model has been trained on a multi-center routine clinical dataset with a range of sequences and slice thickness[16] and with Dice similarity coefficients similar to trained radiologists, there were 10 interval comparisons rejected due to volume segmentation being deemed inadequate (Figure 1).

## CONCLUSION

In conclusion, the application of DL-based fully automated VS volumetric evaluation during a real-world multidisciplinary meeting preparation allowed greater sensitivity for the classification of VS size change when compared to the current "standard of care" approach with maximum linear dimensions; however, this did not result in a significant change in the subsequent clinical management decisions. The successful implementation and added value of such an AI model suggest that it may usefully augment the radiologist interpretive skills; however, its real benefits will only be achieved when there is a clear impact on patient care.

**Availability of Data and Materials:** The data that support the findings of this study are available on request from the corresponding author.

**Ethics Committee Approval:** This study was approved by the Ethics Committee of North of Scotland (approval no.: 22/NS/0160 (AIMBraTS); date: June 12, 2022).

**Informed Consent:** The Ethics Committee of North of Scotland waived the requirement for consent.

**Peer-review:** Externally peer-reviewed.

**Author Contributions:** Concept – S.C., J.S.; Design – S.C., N.W., J.S.; Supervision – J.S.; Resources – J.S., N.W., T.V.; Materials – J.S., T.V.; Data Collection and/or Processing – S.C., N.W., A.O., R.B.; Analysis and/or Interpretation – S.C., N.W.; Literature Search – S.C.; Writing – S.C.; Critical Review – N.W., A.O., R.B., T.V., J.S.

**Declaration of Interests:** The authors have no conflicts of interest to declare.

**Funding:** This work is supported by core funding from Wellcome / EPSRC [WT203148/Z/16/Z; NS/A000049/1] and an MRC project grant (MC/PC/180520). N. Wijethilake is supported by the UK Medical Research Council [MR/N013700/1] and the King's College London MRC Doctoral Training Partnership in Biomedical Sciences.

## REFERENCES

1. Marinelli JP, Beeler CJ, Carlson ML, Caye-Thomasen P, Spear SA, Erbele ID. Global incidence of sporadic vestibular schwannoma: A systematic review. *Otolaryngol Head Neck Surg*. 2022;167(2):209-214. [CrossRef]
2. Shapey J, Barkas K, Connor S, et al. A standardised pathway for the surveillance of stable vestibular schwannoma. *Ann R Coll Surg Engl*. 2018;100(3):216-220. [CrossRef]
3. Li D, Tsimpas A, Germanwala AV. Analysis of vestibular schwannoma size: A literature review on consistency with measurement techniques. *Clin Neurol Neurosurg*. 2015;138:72-77. [CrossRef]
4. Kania R, Vérillaud B, Camous D, et al. EAONO position statement on Vestibular schwannoma: imaging Assessment Question: how should growth of Vestibular schwannoma be defined? *J Int Adv Otol*. 2018;14(1):90-94. [CrossRef]
5. Cross JJ, Baguley DM, Antoun NM, Moffat DA, Prevost AT. Reproducibility of volume measurements of vestibular schwannomas - a preliminary study. *Clin Otolaryngol*. 2006;31(2):123-129. [CrossRef]
6. Harris GJ, Plotkin SR, Maccollin M, et al. Three-dimensional volumetrics for tracking vestibular schwannoma growth in neurofibromatosis type II. *Neurosurgery*. 2008;62(6):1314-9; discussion 1319. [CrossRef]
7. MacKeith S, Das T, Graves M, et al. A comparison of semi-automated volumetric vs linear measurement of small vestibular schwannomas. *Eur Arch Otorhinolaryngol*. 2018;275(4):867-874. [CrossRef]
8. 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(1065):20160110. [CrossRef]
9. Selleck AM, Rodriguez JD, Brown KD. Vestibular schwannoma measurements-is volumetric analysis clinically necessary? *Otol Neurotol*. 2021;42(6):906-911. [CrossRef]
10. van de Langenberg R, de Bondt BJ, Nelemans PJ, Baumert BG, Stokroos RJ. Follow-up assessment of vestibular schwannomas: volume quantification versus two-dimensional measurements. *Neuroradiology*. 2009;51(8):517-524. [CrossRef]
11. Varughese JK, Wentzel-Larsen T, Vassbotn F, Moen G, Lund-Johansen M. Analysis of vestibular schwannoma size in multiple dimensions: a comparative cohort study of different measurement techniques. *Clin Otolaryngol*. 2010;35(2):97-103. [CrossRef]
12. Vokurka EA, Herwadkar A, Thacker NA, Ramsden RT, Jackson A. Using Bayesian tissue classification to improve the accuracy of vestibular schwannoma volume and growth measurement. *AJNR Am J Neuroradiol*. 2002;23(3):459-467.
13. Walz PC, Bush ML, Robinett Z, Kirsch CFE, Welling DB. Three-dimensional segmented volumetric analysis of sporadic vestibular schwannomas: comparison of segmented and linear measurements. *Otolaryngol Head Neck Surg*. 2012;147(4):737-743. [CrossRef]
14. McGrath H, Li P, Dorent R, et al. Manual segmentation versus semi-automated segmentation for quantifying vestibular schwannoma volume on MRI. *Int J Comput Assist Radiol Surg*. 2020;15(9):1445-1455. [CrossRef]
15. Cass ND, Lindquist NR, Zhu Q, Li H, Oguz I, Tawfik KO. Machine learning for automated calculation of vestibular schwannoma volumes. *Otol Neurotol*. 2022;43(10):1252-1256. [CrossRef]
16. Kujawa A, Dorent R, Connor S, et al. Deep learning for automatic segmentation of vestibular schwannoma: A retrospective study from multi-centre routine MRI. *Front Comput Neurosci*. 2024;18:1365727. [CrossRef]
17. Neve OM, Chen Y, Tao Q, et al. Fully automated 3D vestibular schwannoma segmentation with and without gadolinium-based contrast material: A multicenter, multivendor study. *Radiol Artif Intell*. 2022;4(4):e210300. [CrossRef]
18. Shapey J, Wang G, Dorent R, et al. An artificial intelligence framework for automatic segmentation and volumetry of vestibular schwannomas from contrast-enhanced T1-weighted and high-resolution T2-weighted MRI. *J Neurosurg*. 2019;6:1-9.
19. Suresh K, Luo G, Bartholomew RA, et al. An external validation study for automated segmentation of vestibular schwannoma. *Otol Neurotol*. 2024;45(3):e193-e197. [CrossRef]
20. Wang H, Qu T, Bernstein K, Barbee D, Kondziolka D. Automatic segmentation of vestibular schwannomas from $T_1$-weighted MRI with a deep neural network. *Radiat Oncol*. 2023;18(1):78. [CrossRef]
21. Walraven JEW, Verhoeven RHA, Meulen RV, et al. Facilitators and barriers to conducting an efficient, competent and high-quality oncological multidisciplinary team meeting. *BMJ Open Qual*. 2023;12(1):e002130. [CrossRef]
22. Brady A, Laoide RÓ, McCarthy P, McDermott R. Discrepancy and error in radiology: concepts, causes and consequences. *Ulster Med J*. 2012;81(1):3-9.
23. Wijethilake N, Connor S, Oviedova A, et al. Artificial intelligence for personalized management of vestibular schwannoma: A clinical implementation study within a multidisciplinary decision-making environment. *medRxiv*. 2023. [CrossRef]
24. Wijethilake N, Connor S, Oviedova A, Kujawa A, Burger R, Vercauteren T, Shapey J. A. A clinical guideline driven automated linear feature extraction for vestibular schwannoma. Medical imaging: Image Processing.2024; 12926: 617-624.
25. Di Ieva A. AI-augmented multidisciplinary teams: hype or hope? *Lancet*. 2019;394(10211):1801. [CrossRef]
26. Martin M, Kristeleit H, Ruta D, et al. Augmentation of a multidisciplinary team meeting with a clinical decision support system to triage breast cancer patients in the United Kingdom. *Future Medicine AI*. 2023;1. [CrossRef]
27. Mello-Thoms C, Mello CAB. Clinical applications of artificial intelligence in radiology. *Br J Radiol*. 2023;96(1150):20221031. [CrossRef]

28. Lin L, Dou Q, Jin Y-M, et al. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. *Radiology*. 2019;291(3):677-686. [CrossRef]
29. de Boer NP, Böhringer S, Koot RW, et al. A prediction model for recurrence after translabyrinthine surgery for vestibular schwannoma: toward personalized postoperative surveillance. *Eur Arch Otorhinolaryngol*. 2022;279(6):2905-2913. [CrossRef]
30. Langenhuizen PPJH, Zinger S, Leenstra S, et al. Radiomics-based prediction of long-term treatment response of vestibular schwannomas following stereotactic radiosurgery. *Otol Neurotol*. 2020;41(10):e1321-e1327. [CrossRef]
31. Kleijwegt MC, van der Mey AG, Wiggers-deBruine FT, Malessy MJ, van Osch MJ. Perfusion magnetic resonance imaging provides additional information as compared to anatomical imaging for decision-making in vestibular schwannoma. *Eur J Radiol Open*. 2016;3(3):127-133. [CrossRef]
32. Rong Y, Chen Q, Fu Y, et al. NRG oncology assessment of artificial intelligence deep learning-based auto-segmentation for radiation therapy: current developments, clinical considerations, and future directions. *Int J Radiat Oncol Biol Phys*. 2024;119(1):261-280. [CrossRef]
33. Galsgaard A, Doorschodt T, Holten AL, Müller FC, Ploug Boesen M, Maas M. Artificial intelligence and multidisciplinary team meetings; a communication challenge for radiologists' sense of agency and position as spider in a web? *Eur J Radiol*. 2022;155:110231. [CrossRef]
34. Łajczak P, Matyja J, Jóźwik K, Nawrat Z. Accuracy of vestibular schwannoma segmentation using deep learning models - a systematic review & meta-analysis. *Neuroradiology*. 2024. [CrossRef]