Published in final edited form as:

Otol Neurotol. 2018 June ; 39(5): 632–638. doi:10.1097/MAO.0000000000001781.

Improvement in patient-reported hearing after treatment with bevacizumab in people with Neurofibromatosis Type 2

Victoria Huang, BS¹, Amanda L. Bergner, MS CGC^{1,2}, Chris Halpin, PhD³, Vanessa L. Merker, BS³, Monica R. Sheridan, BA³, Brigitte Widemann, MD⁴, Jaishri O. Blakeley, MD^{1,*}, and Scott R. Plotkin, MD, PhD^{3,*}

¹Johns Hopkins University, Baltimore, MD

²Columbia University, New York, NY

³Department of Neurology and Cancer Center, Massachusetts General Hospital, Boston, MA

⁴National Cancer Institute, Bethesda, MD

Abstract

Objective—Assess patient-reported outcomes (PRO) for hearing and tinnitus relative to clinical hearing assessment in people with neurofibromatosis 2 (NF2) associated hearing loss.

Study Design—Prospective, open label, phase-II clinical trial with PRO administered pre-, post- and after treatment.

Setting—Three tertiary referral centers.

Patients—Fourteen patients with NF2, median age of 30 years (range, 14–79 years) and progressive hearing loss (median baseline word recognition score, 60%; range 13% to 82%). Half of these patients achieved objective hearing response (word recognition score improved beyond the 95% critical difference versus baseline).

Intervention—Bevacizumab 7.5 mg/kg was administered every 3 weeks for 48 weeks, followed by surveillance for 24 weeks off-drug.

Main Outcome Measures—Speech, Spatial and Qualities of Hearing Scale (SSQ) and Tinnitus Reaction Questionnaire (TRQ) to assess hearing difficulties in life situations and tinnitus related distress.

Results—Patient-reported speech understanding and auditory quality improved with bevacizumab treatment and were significantly correlated with word recognition scores, but not pure tone threshold average. There was no change in spatial perception after treatment. Reduction in tinnitus distress after treatment with bevacizumab did not reach statistical significance.

Conclusion—Participants had reductions in hearing difficulty during treatment with bevacizumab, suggesting that patients subjectively experience hearing-related benefit mirroring clinical hearing assessments. We suspect the lack of significant reduction in tinnitus distress is

Corresponding author: Scott R. Plotkin, MD, PhD, Yawkey 9E, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, Telephone: 617-726-3650, Fax: 617-643-2591, splotkin@partners.org.

*Denotes equal contribution

related to small sample size and low intensity of distress in our sample. These data highlight the utility of PRO measures to assess benefits of treatment in the setting of NF2-associated hearing loss.

Keywords

Bevacizumab; Neurofibromatosis 2; Vestibular Schwannoma; Patient Reported Outcome Measures; Hearing; Tinnitus

INTRODUCTION

Neurofibromatosis type 2 (NF2) is a tumor predisposition syndrome characterized by the occurrence of bilateral vestibular schwannomas (1). Most patients with NF2 have progressive bilateral hearing loss and tinnitus (2). Retrospective studies of patients treated with bevacizumab, a monoclonal antibody that binds vascular endothelial growth factor (VEGFR), showed an improvement in hearing on audiologic testing in some patients with NF2 (3, 4). These findings were confirmed in a prospective, multi-institution, open label phase II clinical trial which documented objective improvement in word recognition scores for 12 of 19 evaluable ears in 7 of 14 patients, with 5 of 14 patients having durable, statistically significant improvement (5).

To date, patient reported outcomes (PRO) for hearing have not been assessed in any studies with bevacizumab. The US Food and Drug Administration and others have highlighted the importance of assessing patient reported outcomes in addition to clinical outcomes in clinical trials (6, 7). Patient reported outcome measures (PROMs) can contribute to the assessment of net clinical benefit of an intervention and subsequent measurement of cost-effectiveness (8,9). For otology and neurotology, PROMs have been shown to help capture the real-world impact of disease symptoms and interventions on participants' lives, complementing standard clinical measures such as audiology (10,11). While standard audiologic evaluation involves measurement of hearing under a single, controlled condition, hearing-related PROMs can add complementary information about perception in the varied conditions experienced by patients in their daily lives (12).

METHODS

Fourteen participants with NF2 were enrolled in a phase II trial assessing the effect of bevacizumab on vestibular schwannoma associated hearing loss (5). The target tumor (ear) for each participant was the vestibular schwannoma causing active, progressive hearing loss. Each participant had pure tone threshold testing by audiologists using standard clinical procedures (13). The individual thresholds were combined across frequency (500, 1000, 2000 and 4000 Hz) into a pure tone average (PTA) in order to capture the overall effect of the patient's hearing loss on audibility, particularly for speech (14). Patients' word recognition ability was also assessed using 100-word recorded monosyllable lists (CID W22) delivered at an intensity calculated for maximum performance. Because these participants had vestibular schwannomas, word recognition was checked at the standard maximum performance level, and again at a level 10 dB lower. The greater of those two word recognition scores (WRS) is reported here in order to ensure that the true maximum

score was reported, unaffected by the possible “rollover” (i.e. scores paradoxically reduced with increased intensity) sometimes seen in cases with vestibular schwannoma (15).

The hearing PROM used was the Speech, Spatial and Qualities of Hearing (SSQ) scale (12). This tool consists of 49 questions divided into three subscales addressing difficulty with speech understanding (14 questions), spatial location of sounds (17 questions) and the qualities of sounds (18 questions). The speech domain covers a range of real-world contexts and difficulty levels, including variations on the number of speakers in a conversation, speakers’ visibility, and types of background noise. The spatial domain consists of the ability to judge the direction, distance, and movement of sounds. The qualities domain includes the ability to segregate two simultaneous sounds as distinct entities, the ability to recognize familiar sounds, the clarity and naturalness of speech and other noises, and listening effort. Responses are collected using an 11-point Likert scale (0 to 10) describing the patient’s self-perceived ability in specified scenarios with the anchor points “Perfectly” (= 10) and “Not at all” (= 0) at either end. For each subscale, we used the mean of participants’ responses to produce three domain-specific summary scores (also on a 0 to 10 scale with higher scores indicating better hearing). The SSQ has been validated regarding modes of administration (16), relation to other hearing variables, and the internal consistency of items (as measured using Cronbach’s alpha) (17), but has not been applied to clinical trials of chemotherapy for vestibular schwannoma.

Psychological distress caused by tinnitus was evaluated using the Tinnitus Reaction Questionnaire (TRQ) (18). This tool evaluates patients’ negative views of their tinnitus, but does not capture the nature of the percept itself (e.g. pitch, loudness, location). The TRQ consists of a series of 26 statements such as “my tinnitus has made me unhappy” and the patient is asked to respond using a 5-point Likert scale ranging from 0 (not at all) to 4 (almost all of the time). The responses are summed, resulting in a range of 0–104 with higher scores indicating more distress related to tinnitus. We considered participants to be substantially bothered by tinnitus when their scores were above 32.5, which was the mean of scores in the original TRQ publication (18). The SSQ and TRQ were self-administered via paper and pencil at baseline, week 24, week 48 (on treatment) and after 24 weeks off treatment (week 72). Responses at week 24 were not included in this report since that administration had a higher rate of missing data (28.6% SSQ; 14.3% TRQ) and did not represent a study milestone for clinical assessments. Participant completion of the SSQ and TRQ was 100% at baseline, 93% (one missing case) at week 48 and 93% at week 72.

Differences in groups were assessed using the Mann-Whitney U test. Pair-wise changes across time points were assessed using the Wilcoxon Signed Rank test and the strength of correlations were assessed using Spearman’s Rho. Effect sizes (r) were calculated and qualitatively described following Cohen’s (1998) criteria of 0.1 = small effect, 0.3 = medium effect, 0.5 = large effect (19, 20). Most comparisons are assessed conservatively using two-tailed analysis except for when assessing the difference in spatial ability between binaural and monaural cases; prior evidence indicates that binaural patients have better spatial recognition, so a one-tailed analysis was used and noted (21).

RESULTS

Audiologic and radiographic results of treatment with bevacizumab for this trial have been published previously (5). Participants' ages ranged from 14–79 years (median 30) and 71% (10/14) were female. One participant was removed from treatment early due to toxicity; this participant did not complete the week 48 scales, but did complete the week 72 scales. Another participant failed to complete the week 72 (off treatment) scales.

Five participants had two hearing ears and nine had one hearing ear, with the contralateral ear deafened from prior surgical resection of a vestibular schwannoma. Seven of the participants (50%) showed at least one instance of a hearing response to bevacizumab, defined as an improvement in word recognition exceeding the 95% critical difference versus the participant's baseline score. The median PTA at baseline was 51.25 dB HL (range 17.50 – 93.75) and the median word recognition score was 60 percent correct (range 13% - 82%). The clinical hearing status of all participants at baseline is shown in figure 1.

SSQ Questionnaire results at baseline

The initial scores for all participants are shown in figure 2. The median score for speech was 3.15 (on the 0–10 Likert scale; range 1.0 to 6.2). The median score for spatial effects was 3.25 (range 0.3 to 8.6), and the median score for auditory qualities was 4.7 (range 0.1 to 8.9). The perception of spatial localization presents a special consideration in this population due to the large proportion of monaural participants. Nine (64%) of the 14 participants had only one hearing ear, and this is expected to substantially limit auditory localization ability (17). The perception of auditory space and sound location was higher for binaural subjects compared with monaural subjects with scores of 7.2 vs. 3.0 (Mann-Whitney U; $p = 0.0415$ [one-tailed]).

Relationship between patient-reported and clinical hearing scores

The relationship between patient-reported hearing function and their clinical hearing scores (WRS and PTA in the target ear) at baseline is shown in figure 3. There was a statistically significant correlation of participants' SSQ speech score with their clinical word recognition score ($\rho = 0.75$; $p = 0.0018$) but not with their pure tone average ($\rho = -0.40$, $p = 0.159$). There was also a statistically significant correlation of participants' SSQ sound qualities score with their word recognition score ($\rho = 0.60$; $p = 0.023$), but not with their pure tone average ($\rho = -0.22$, $p = 0.459$). Participants' SSQ spatial scores were not significantly correlated with either WRS ($\rho = 0.47$, $p = 0.09$) or PTA ($\rho = -0.17$, $p = 0.55$).

Change in patient reported hearing after treatment with bevacizumab

Figure 4 shows the median SSQ results at baseline, week 48 (end of treatment) and at week 72 (24 weeks off treatment). From baseline to week 48, participants' reported ability to understand speech improved significantly (median change +1.8, $p = 0.0046$; $r = 0.556$). The perception of the quality of sounds also improved significantly on treatment (median change +1.3, $p = 0.0232$; $r = 0.446$). In contrast, auditory spatial ability did not significantly improve on treatment (median change +1.3, $p = 0.1336$). Due to low sample size, we could not statistically assess change of auditory spatial ability in only binaural patients. Statistical

comparison of week 48 results to week 72 showed no significant differences (speech: $p = 0.101$, quality: $p = 0.271$; spatial: $p = 0.332$). The comparison of baseline results to the off-drug point also did not reach statistical significance (speech: $p = 0.051$, quality: $p = 0.123$, spatial: $p = 0.271$).

Half of patients had significantly improved word recognition scores in the target ear with bevacizumab treatment. To assess if clinical hearing responders also had larger improvements in self-reported hearing than non-responders, we compared the change in SSQ scores in these two groups from baseline to end of treatment (week 48). The median changes for responders was greater than for non-responders, respectively: +2.05 vs. +1.0 for speech, +2.3 vs. +0.9 for spatial location, and +1.45 vs. +0.1 for auditory qualities. However, none of these comparisons were statistically significant (Mann-Whitney U; $p = 0.430$, 0.552 and 0.617, respectively.).

Tinnitus distress

Figure 5 shows tinnitus questionnaire scores at baseline, week 48 (end of treatment) and week 72 (24 weeks off treatment). Median TRQ score at baseline was 17.5 (on a 104 point scale), with 5 of 14 (36%) participants substantially bothered by tinnitus at baseline. Two participants reported zero tinnitus distress across all time points. Self-reported tinnitus distress did not significantly correlate with baseline word recognition score ($\rho = 0.05$; $p = 0.867$), or with their PTA ($\rho = -0.23$; $p = 0.434$). Baseline TRQ scores were also not significantly different between binaural and monaural cases (Mann-Whitney U; $p = 0.99$). Self-reported negative reaction to tinnitus did not significantly decrease from baseline to week 48 (median change -7 ; $p = 0.056$) or between baseline and week 72 (median change -5 , $p = 0.12$). Change from week 48 to week 72 (median change: 0 points) could not be statistically evaluated due to low sample size. Tinnitus distress scores were also compared in participants' with and without a hearing response to bevacizumab. The change in tinnitus scores from baseline to the end of treatment (week 48) were -9.5 for responders and -2 for non-responders; this difference was not statistically significant (Mann-Whitney U; $p = 0.6672$).

DISCUSSION

To date, there have been no published studies of PROMs in treatment trials of chemotherapy for NF2. We have shown that both patient-reported ability to understand speech and the perception of auditory qualities showed statistically significant improvements of moderate to large effect size ($r = .556$ and 0.446 , respectively) after treatment with bevacizumab in people with hearing loss related to NF2-associated vestibular schwannomas. This is consistent with the previous finding of significant improvement in clinical hearing results by word recognition (5). The findings with SSQ and TRQ in this study indicate that PRO measures are feasible for studies of NF2 associated hearing loss and that they add context to the clinical evidence of hearing benefit with bevacizumab. PROs are needed in order to capture the patient's perspective on the benefit of therapy and may be used in future studies to better understand the cost-benefit ratio of bevacizumab (8). Importantly, PROs are best used to

complement standard clinical measurements, and cannot be substituted for validated clinical assessments, such as audiologic evaluation.

PTA (threshold) findings are often thought to represent “hearing” with no need to explore the auditory quality and utility represented by word recognition. In reality, PTA only reflects a threshold for a perception of a sound. In order to maximize the frequency (cochlear place) specificity of the test, all other auditory requirements besides the simplest binary detection task are removed. The pure tone threshold, therefore, functions well as a cochlear regional map of healthy and dysfunctional cochlear regions, but does not directly scale the amount or type of cochlear damage. (22, 23) In contrast, WRS measures the capacity of the auditory system to supply enough information for complex tasks. In this study, all of the improvements in the PRO assessment of hearing function correlated with improvement in WRS, but not PTA. Hence, while progressively worsening thresholds (PTA) do limit audibility, the PRO results here support that the standard clinical measure of WRS better reflects the difficulty of the patient in daily activities related to hearing. WRS should thus be considered the optimal metric for objective hearing function in both clinical practice and clinical research for NF2 associated vestibular schwannoma (24).

Thirty-six percent of patients reported substantial distress related to tinnitus at baseline, and TRQ scores were not significantly related to degree of hearing loss by either WRS or PTA. The independence of this effect from WRS or PTA scores is consistent with the clinical observation that the negative experience of tinnitus and severity of hearing loss are often uncoupled in cases with vestibular schwannoma (25). The improvement in tinnitus distress after treatment with bevacizumab treatment was not statistically significant, but our analysis was limited given that two participants reported distress scores of zero across the entire study period. Further evaluation of bevacizumab’s effects on tinnitus distress should focus on individuals with greater baseline impairment or use alternate measures of tinnitus. For example, the Tinnitus Handicap Inventory (THI) might be preferable to the TRQ given our finding of low baseline tinnitus distress and the measure’s recent recommendation for use in clinical trials (26).

A notable limitation of this study is the lack of a control group. A control group is always preferred when assessing subjective reports such as are captured using PROs. This study represents a report of a secondary endpoint from a clinical trial, but future studies designed with PROs as primary endpoints should include a control group. In assessing these results, it is reassuring that only two of the three measures in the SSQ improved. Specifically, there was no improvement in the spatial subscale (Figure 4), which would have been unexpected given that only 5/14 subjects had two hearing ears and spatial localization relies on binaural hearing. This difference in changes on the spatial ability subscale versus the speech and qualities subscales offers some evidence that an indiscriminate placebo effect from open-label treatment is not responsible for the results; however, this can not be conclusively proven without a control group.

The study is also limited by its small sample size. This phase II clinical trial was powered to detect a change in WRS with bevacizumab treatment. An N of 14 was estimated to provide 90% power to detect an alternate response rate on the clinical hearing variables at an alpha

level of $P = 0.05$ (5). All secondary analyses were exploratory and hence, it is unknown what relationships may be found with appropriately powered studies dedicated to the analysis of change in PROMs.

An advantage of the SSQ and TRQ is that they are function-specific PROMs relevant to people with vestibular schwannomas rather than general health-related QOL measures which could be influenced by the myriad of other disease manifestations that affect people with NF2. Larger efforts to determine minimally clinically important differences for these measures in the future would enhance the utility and interpretation of PROM results (27, 28). The agreement between PRO measurements and objective assessment of hearing supports a broader interpretation of the previously reported hearing improvement with bevacizumab, namely that hearing improvement demonstrated in the standardized conditions of audiology testing is also apparent across a multitude of real-world hearing contexts important to patients in their daily lives.

Acknowledgments

Conflict of Interest and Sources of Funding:

Amanda Bergner has received a salary from Ambry Genetics.

Brigitte Widemann had no conflict of interest to report. Her research is supported by the Intramural Research program of the Center for Cancer Research, NCI

Vanessa Merker receives funding from the Program in Cancer Outcomes Research Training (NCI R25CA92203) and a Young Investigator's Award from the Children's Tumor Foundation.

Jaishri Blakeley receives research funding from NIH, the Children's Tumor Foundation, the Galloway Foundation, the Neurofibromatosis Therapeutic Acceleration Program and private compensation as a consultant for Abbvie Inc.

Scott Plotkin receives research funding from Children's Tumor Foundation, NIH, and the Department of Defense Neurofibromatosis Research Program. He has consulted for Novartis.

For the remaining, authors none were declared

REFERENCES

1. Asthagiri A, Parry D, Butman J, et al.: Neurofibromatosis type 2. *Lancet* 2009; 373:1974–1986. [PubMed: 19476995]
2. Plotkin S, Merker V, Muzikansky A, et al.: Natural history of vestibular schwannoma growth and hearing decline in newly diagnosed neurofibromatosis type 2 patients. *Otol Neurotol* 2014; 35:50–56.
3. Plotkin S, Stemmer-Rachamimov A, Halpin C, Tyrell A, Padera T, Sorenson G, Jain R and Di Tomaso E Hearing improvement after bevacizumab in patients with neurofibromatosis 2. *New Eng J Med* 2009; 361:358–367. [PubMed: 19587327]
4. Plotkin S, Merker V, Halpin C, McKenna M and Barker F Bevacizumab for Progressive Vestibular Schwannoma in Neurofibromatosis Type 2: A Retrospective Review of 31 patients *Otol Neurotol* 2012; 33(6):1046–52. [PubMed: 22805104]
5. Blakeley J, Ye X, Duda D, Halpin C, Bergner A, Muzikansky A, Gerstner E, Merker V, Fayad L, Ahlawat S, Jacobs M, Jain R, Dombi E, Widemann B and Plotkin S Efficacy and biomarker study of bevacizumab for hearing loss due to neurofibromatosis type 2 associated vestibular schwannomas. *J Clin Onc* 2016; 34(14):1669–1675.
6. Speight J and Barendse S FDA guidance on patient reported outcomes. *BMJ* 2010; 340:2921

7. Coons SJ, Kothari S, Monz BU, Burke LB. The patient-reported outcome (PRO) consortium: filling measurement gaps for PRO end points to support labeling claims. *Clin Pharmacol Ther.* 2011 11;90(5):743–8. [PubMed: 21993428]
8. Arpinelli F, Bamfi F. The FDA guidance for industry on PROs: the point of view of a pharmaceutical company. *Health Qual Life Outcomes.* 2006 10 31;4:85. [PubMed: 17076891]
9. Armstrong TS. Measuring clinical benefit: use of patient-reported outcomes (PRO) in primary brain tumor clinical trials. *Curr Oncol Rep.* 2013 2;15(1):27–32. [PubMed: 23054936]
10. Powell J, Powell S, Robson A. A systematic review of patient-reported outcome measures in paediatric otolaryngology. *J Laryngol Otol.* 2018 1;132(1):2–7. [PubMed: 29224575]
11. Shaffer BT, Cohen MS, Bigelow DC, Ruckenstein MJ. Validation of a disease-specific quality-of-life instrument for acoustic neuroma: the Penn Acoustic Neuroma Quality-of-Life Scale. *Laryngoscope.* 2010 8;120(8):1646–54. [PubMed: 20641085]
12. Gatehouse S and Noble W The speech, spatial and qualities of hearing scale (SSQ). *Int J Aud* 2004; 43:85–99.
13. American National Standards Institute. S3.21 Methods for manual pure tone threshold audiometry. New York: Author, 1978.
14. Halpin C Measuring audiometric outcomes In: Shin J, Hartnick C and Randolph G, eds Evidence Based Otolaryngology. New York: Springer 2008; 227–238.
15. Bess F, Josey A, and Humes L Performance intensity functions in cochlear and eighth nerve disorders. *Am J Otol* 1979; 7, 1(1): 27–31. [PubMed: 554465]
16. Singh G and Pichora-Fuller M Older adults' performance on the speech, spatial and qualities of hearing scale (SSQ): Test-retest reliability and a comparison of interview and self administration methods. *Int J Aud* 2010; 49:733–740.
17. Akeroyd M, Guy F, Harrison D and Suller S A factor analysis of the SSQ (Speech, Spatial, and Qualities of Hearing Scale) *Intl J Aud* 2014; 53: 101–114.
18. Wilson P, Henry J, Bowen M and Haralambos G Tinnitus reaction Questionnaire: Psychometric properties of a measure of distress associated with tinnitus. *J Sp Hrg Res* 1991; 34:197–201.
19. Cohen J Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic, 1988.
20. Rosenthal R Parametric measures of effect size In Cooper H, Hedges LV, eds, The handbook of research synthesis. New York: Russell Sage Foundation, 1994; 231–244.
21. Hausler R, Colburn H, Marr E. Sound localization in subjects with impaired hearing. *Acta Oto Laryngol* 1983;[suppl 400]: Monograph.
22. Halpin C The tuning curve in clinical audiology. *Am J Aud* 2002; 11:56–64.
23. Halpin C and Rauch S Clinical implications of a damaged cochlea: pure tone thresholds versus information carrying capacity *Otolaryng - HNS* 2009; 140(4):473–476.
24. Plotkin S, Ardern-Holmes S, Barker F II, Blakeley J, Evans T, Ferner R, Hadlock T and Halpin C Hearing and facial function outcomes for neurofibromatosis 2 clinical trials. *Neurology* 2013; 81: S25–S32. [PubMed: 24249803]
25. Stouffer J, Tyler R, Kileny P and Dalzell L Tinnitus as a function of duration and etiology: counseling implications. *Am J Otol* 1991;12:188–194.
26. Lamdgrebe M, Azevedo A, Baugley D, et al. Methodological aspects of clinical trials in tinnitus: a proposal for an international standard. *J Psychosom Res.* 2012; 73(2):112–21. [PubMed: 22789414]
27. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J.* 2007 Sep-Oct;7(5): 541–6. [PubMed: 17448732]
28. Carlson ML, Tveiten ØV, Yost KJ, Lohse CM, Lund-Johansen M, Link MJ. The Minimal Clinically Important Difference in Vestibular Schwannoma Quality-of-Life Assessment: An Important Step beyond $P < .05$. *Otolaryngol Head Neck Surg.* 2015 8;153(2):202–8. [PubMed: 26038393]

		Maximum Word Recognition %									
		100-90	89-80	79-70	69-60	59-50	49-40	39-30	29-20	19-10	9-0
Pure Tone Threshold Average (dB HL)	0-10										
	11-20								1		
	21-30		1	1							
	31-40			1							
	41-50		1					2			
	51-60			1	1						
	61-70			1		1					
	71-80									1	
	81-90							1			
	91-100							1			

Figure 1. Baseline Hearing.

Fourteen cases comprising the NF2 cohort are shown before treatment with bevacizumab. These results are presented for patients' target ears using the AAO recommended format (<http://hearingoutcomes.stanford.edu>) with maximum word recognition (percent correct of monosyllables) along the horizontal axis and pure tone threshold average (dB HL, 500, 1000, 2000 and 4000 Hz) down the vertical axis.

Figure 2. Baseline Speech, Spatial and Qualities Hearing Scale scores.

The vertical axis shows SSQ subscale scores, where higher scores represent better responses (less hearing difficulty) and lower scores represent worse responses (more hearing difficulty). The horizontal axis has three columns per study participant, one for each of the SSQ subscales.

Figure 3. SSQ Speech and Qualities subscale results versus hearing at baseline.

The vertical axes show participants' baseline SSQ subscale scores. In panels (a) and (b), the horizontal axes show the percent correct of monosyllable scores for each participant. There was a significant correlation of the speech subscale scores with word recognition (Spearman's ρ ; $P=0.0018$), as well as between the qualities subscale and word recognition scores ($P=0.023$). In the lower panels (c) and (d), the horizontal axis shows the pure tone threshold average the participants. Correlation of both speech and quality subscales with the baseline pure tone average did not reach statistical significance.

Figure 4. SSQ results before, during, and after treatment with bevacizumab

There are three panels, one for each of the SSQ subscales. The vertical axes show the SSQ scores for each case, wherein higher scores represent better outcomes (less hearing difficulty). The horizontal axis marks the study timepoints at baseline, the end of treatment (week 48) and the off-study evaluation (week 72). Brackets with an asterisk indicate a statistically significant improvement at the $P < 0.05$ level.

Figure 5. Tinnitus Reaction Questionnaire results before, during, and after treatment with bevacizumab

TRQ scores for all cases are shown at baseline, week 48 and off-study (week 72). The vertical axis is the TRQ score. The bracket with an asterisk indicates a statistically significant improvement (Wilcoxon Signed Rank Test; $P=0.028$ [one-tailed]).