Hearing and facial function outcomes for neurofibromatosis 2 clinical trials

Scott R. Plotkin, MD
PhD
Simone L. Ardern-
Holmes, MD
Fred G. Barker II, MD
Jaishri O. Blakeley, MD
D. Gareth Evans, MD
Rosalie E. Ferner, MD
Tessa A. Hadlock, MD
Chris Halpin, PhD
For the REiNS
International
Collaboration

Correspondence to
Dr. Plotkin:
splotkin@partners.org

ABSTRACT

Objectives: Vestibular schwannomas are the hallmark of neurofibromatosis 2 (NF2), occurring in >95% of patients. These tumors develop on the vestibulocochlear nerve and are associated with significant morbidity due to hearing loss, tinnitus, imbalance, facial weakness, and risk of early mortality from brainstem compression. Although hearing loss and facial weakness have been identified as important functional outcomes for patients with NF2, there is a lack of consensus regarding appropriate endpoints in clinical trials.

Methods: The functional outcomes group reviewed existing endpoints for hearing and facial function and developed consensus recommendations for response evaluation in NF2 clinical trials.

Results: For hearing endpoints, the functional group endorsed the use of maximum word recognition score as a primary endpoint, with the 95% critical difference as primary hearing outcomes. The group recommended use of the scaled measurement of improvement in lip excursion (SMILE) system for studies of facial function.

Conclusions: These recommendations are intended to provide researchers with a common set of endpoints for use in clinical trials of patients with NF2. The use of common endpoints should improve the quality of clinical trials and foster comparison among studies for hearing loss and facial weakness. *Neurology*® 2013;81 (Suppl 1):S25–S32

GLOSSARY

AAO-HNS = American Academy of Otolaryngology-Head and Neck Surgery; **HB** = House-Brackmann; **HR** = hearing response; **NF2** = neurofibromatosis 2; **PTA** = pure-tone average; **REiNS** = Response Evaluation in Neurofibromatosis and Schwannomatosis; **SH** = stable hearing; **SMILE** = scaled measurement of improvement in lip excursion; **VS** = vestibular schwannoma; **WRS** = word recognition score.

Vestibular schwannomas (VS) are the hallmark of neurofibromatosis 2 (NF2), occurring in >95% of patients. VS develop on the vestibulocochlear nerve and cause significant morbidity due to hearing loss, tinnitus, imbalance, and risk of early mortality from brainstem compression. Hearing loss develops over time in patients with NF2, and these declines are associated with gradual VS growth.[1] Facial weakness is a common sequela of surgery to remove VS and may also occur as a late manifestation of large VS, secondary to facial nerve schwannomas, or as an isolated mononeuropathy.[1] Facial weakness causes articulation difficulties, oral incompetence, lack of ability to smile and express emotions, and inability to blink and protect the cornea, which can lead to vision loss. Facial weakness and hearing loss adversely affect quality of life[2–4] and as such represent important functional outcomes in clinical trials.

NATURAL HISTORY OF HEARING LOSS AND FACIAL WEAKNESS IN NF2 The natural history of hearing loss in NF2 has been studied in the prospective multicenter NF2 Natural History Study sponsored by the Department of Defense Consortium.[5] In a cohort of 63 patients (108 VS) within 2 years of the diagnosis of NF2, 27% of the ears experienced a significant loss in pure-tone average, defined as a change of 5 dB compared with baseline.[5] The natural history of facial weakness in NF2 has not been well studied. The best data

come from large retrospective studies of facial function after resection of NF2-related VS. These studies suggest that preoperative tumor size does not correlate with facial function after middle fossa approach or translabyrinthine approach.[6,7]

Consensus endpoints to study hearing loss and facial weakness in the setting of NF2 are necessary. Currently, there is no standard definition of hearing loss or facial weakness that is used uniformly across clinical trials. The lack of consensus makes it difficult to compare results between trials or between clinical reports of surgical outcomes. The NF community would benefit from the use of consensus definitions in future trials, particularly those that will be presented to national regulatory bodies as evidence for drug approval.

MEASUREMENT OF HEARING LOSS AND FACIAL WEAKNESS Hearing is monitored in clinical practice by measuring pure-tone thresholds and word recognition scores (WRS), also known as speech recognition scores or speech discrimination scores. Pure-tone thresholds measure the minimum sound level that an ear can perceive. Thresholds are typically measured at octaves and half-octaves from 250 Hz to 8,000 Hz. An average of thresholds at 500, 1,000, 2,000, and 4,000 Hz (pure-tone average [PTA]) is a recommended standard outcome measure for reporting in cases of VS.[8] WRS measure the ability of patients to meaningfully interpret sounds. WRS are measured as the percentage of monosyllables patients can correctly identify when they are presented at a fully audible level. Paradoxically, WRS sometimes decrease in individuals with VS when the words are presented at high volume (rollover effect).[9] Standard audiologic techniques have been developed to avoid potential pitfalls in testing patients with retrocochlear lesions that exhibit rollover. Patients can be classified into categories of hearing based on American Academy of Otolaryngology-Head and Neck Surgery (AAO-HNS) schemes: class A (PTA ≤ 30, WRS >70%), class B (PTA >30 to ≤ 50, WRS ≥ 50%), class C (PTA >50, WRS ≥ 50%), and class D (PTA any level, WRS <50%). Class A and B hearing are termed "serviceable"; class C and D hearing are termed "nonserviceable."

In clinical practice, facial function is monitored using the House-Brackmann (HB) scale,[10] which grades facial function in 6 steps from normal (HB I) to total paralysis (HB VI) (table 1). In 1985, the AAO-HNS adopted the HB as a universal standard. Over the years, the HB grading scale has been updated.[11] Other scales, such as Burres-Fisch, Nottingham, Sunnybrook, and Yanagihara, have also been used to grade facial function.[12]

METHODS The Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) functional outcomes group was formed to address the lack of consensus about appropriate functional endpoints to use in NF clinical trials. The functional outcomes group has 10–15 active participants that meet monthly for teleconferences and present recommendations biannually to the full REiNS International Collaboration group. The group consists of professionals from various disciplines involved in NF care, including neurologists, geneticists, otolaryngologists, audiologists, oncologists, therapists, radiologists, and statisticians. The group has representation from the US, Europe, and Australia.

Selection of functional endpoints. During the first meeting, the functional endpoint group agreed to discuss 4 functional endpoints that were relevant for patients with NF1, NF2, and schwannomatosis as part of a clinical trial: 1) hearing function, 2) facial function, 3) walking function, and 4) pulmonary function. We report the first 2 endpoints assessed by the functional endpoints working group.

Rating of functional endpoints. The functional outcomes group modified the standardized form used for patient-reported outcomes (see Wolters et al., this supplement). Using this form, the group reviewed outcome measures for hearing function and

Table 1		Grading of facial strength using the House-Brackmann scale[10]
Grade	Description	Characteristics
I	Normal	Normal facial function in all areas
II	Mild dysfunction	Gross: slight weakness noticeable on close inspection; may have very slight synkinesis
		At rest: normal symmetry and tone
		Motion
		Forehead: moderate to good function
		Eye: complete closure with minimum effort
		Mouth: slight asymmetry
III	Moderate dysfunction	Gross: obvious but not disfiguring difference between 2 sides; noticeable but not severe synkinesis, contracture, and/or hemifacial spasm
		At rest: normal symmetry and tone
		Motion
		Forehead: slight to moderate movement
		Eye: complete closure with effort
		Mouth: slightly weak with maximum effort
IV	Moderately severe dysfunction	Gross: obvious weakness and/or disfiguring asymmetry
		At rest: normal symmetry and tone
		Motion
		Forehead: none
		Eye: incomplete closure
		Mouth: asymmetric with maximum effort
V	Severe dysfunction	Gross: only barely perceptible motion
		At rest: asymmetry
		Motion
		Forehead: none
		Eye: incomplete closure
		Mouth: slight movement
VI	Total paralysis	No movement

facial function. For hearing function, the review included use of PTA, AAO-HNS hearing classification,[13] the Gardner-Robertson hearing classification,[14] and the 95% critical difference for word recognition.[15] For facial function, the group reviewed use of the HB scale[11,16] and the scaled measurement of improvement in lip excursion (SMILE) analysis.[17]

RESULTS For hearing endpoints, the functional group chose to endorse the use of maximum WRS as a primary endpoint, with the 95% critical difference for primary hearing outcomes[15] and the SMILE system for facial outcomes.[18] These outcomes are discussed below.

Recommended primary outcome for hearing: Maximum WRS. Hearing in patients with NF2 is measured using an audiogram that includes measurement of pure-tone thresholds and WRS. Word recognition is the measure most closely associated with daily hearing function since it measures the ability to comprehend speech (rather than "detect" it). If word recognition quality improves, the patient can converse successfully, even if a hearing aid is needed to make sounds sufficiently loud.

Different statistical models have been used in clinical trials to compare 2 word recognition tests (as in a within-subject change with treatment). The use of a fixed criterion for significance (e.g., a difference of 15 percentage points from baseline) is attractive given the simplicity of implementing this rule. However, a fixed criterion could allow certain errors, in that changes that exceed the criterion might not be significant at $p = 0.05$ (false-positive) and changes that are less than the criterion might be significant at $p = 0.05$. If instead the significance level is fixed at 0.05, the criterion for difference varies depending on the baseline WRS. This results in 2 scores, 1 above and 1 below the starting value, which are known as the critical differences at the selected probability level (i.e., $p = 0.05$, table 2).[15]

Feasibility. Audiology is widely available and it takes about 20 minutes to perform testing that includes a 50-word recognition list. Standardization of testing is critical for use in clinical trials. For this reason, the group recommends use of a recorded list of standardized monosyllables.

Patient characteristics. Standard audiometry can be administered reliably to individuals aged 5 years and older.[19] Audiometry is valid for important subgroups such as the elderly and people with intellectual disability.[20] Since the results are analyzed as repeated measures, a participant may be evaluated using English materials if the audiologist decides that the results reasonably reflect comparative (test-to-test) performance. A recommended approach may be to require that participants be "testable in English" as opposed to the narrower criterion of English as a first language.[8] There are also valid tests that can be reported using the same mathematical criteria available in several languages.[21]

Use in published studies. The 95% critical difference has been used as a primary endpoint or as a secondary endpoint in studies of sudden sensorineural hearing loss[15] and in studies of bevacizumab and lapatinib for NF2-related VS.[22–24]

Baseline documentation of "target" and "nontarget" lesions. When 2 functional ears are present at baseline, the target ear and nontarget ear should be identified and recorded at baseline. The goal is to target an ear that lends itself to reproducible repeated measures. Target ears should be selected on the basis of the quality of the hearing and the presence of an ipsilateral VS. Hearing loss in the target ear should be attributed to VS (rather than other causes of hearing loss).

The maximum WRS should be recorded for all ears. To address rollover, an empirical search using at least 2 levels ("high" and "low") should be used to identify maximum WRS at each evaluation. Speech intelligibility should be evaluated for each ear using standard word recognition of monosyllables.[25] All tests should use recorded lists (e.g., CID-W22, Ira Hirsh recording, Q/MASS v 2) from a compact disc[26] rather than spoken word lists since this results in less variability of results. The initial level should be set where maximum performance is expected for each ear. The "high" level will be the level calculated for maximum audibility, as in the standard method for level setting in nontumor cases. The "low" level will be the maximum level minus 10–15 dB as chosen by the audiologist to reflect a level with less chance of rollover. The maximum WRS is defined as the greater WRS determined at the "high" and "low" levels.

Response criteria. Hearing response criteria are defined in reference to the baseline WRS at study initiation (table 2). Hearing response (HR) is defined as an improvement in WRS above the 95% critical difference. Stable hearing (SH) is defined as persistence of WRS within the 95% critical difference. Hearing decline is defined as a decrease in WRS below the 95% critical difference.

Special notes on assessment of hearing. WRS is not a continuous variable: it is a sum of binary responses (correct/incorrect) that ranges from 0% to 100%. The existence of upper and lower boundaries introduces a "ceiling" and "floor" effect for patients with hearing near the top and bottom of the range, respectively. As table 2 shows, patients with WRS ≤ 6% are not eligible for hearing decline since the 95% critical difference includes 0%; similarly, patients with WRS ≥94% are not eligible for hearing improvement since the 95% critical difference includes 100%. In the NF2 Natural History study, the rate of spontaneous increase in WRS

Table 2	Hearing response guidelines		
Baseline word recognition score (%)	95% critical difference (%)	Hearing response (%)	Progressive hearing loss (%)
0	0-4	≥6	n/a
2	0-10	≥12	n/a
4	0-14	≥16	n/a
6	2-18	≥20	0
8	2-22	≥24	0
10	2-24	≥26	0
12	4-26	≥28	≤2
14	4-30	≥32	≤2
16	6-32	≥34	≤4
18	6-34	≥36	≤4
20	8-36	≥38	≤6
22	8-40	≥42	≤6
24	10-42	≥44	≤8
26	12-44	≥46	≤10
28	14-46	≥48	≤12
30	14-48	≥50	≤12
32	16-50	≥52	≤14
34	18-52	≥54	≤16
36	20-54	≥56	≤18
38	22-56	≥58	≤20
40	22-58	≥60	≤20
42	24-60	≥62	≤22
44	26-62	≥64	≤24
46	28-64	≥66	≤26
48	30-66	≥68	≤28
50	32-68	≥70	≤30
52	34-70	≥72	≤32
54	36-72	≥74	≤34
56	38-74	≥76	≤36
58	40-76	≥78	≤38
60	42-78	≥80	≤40
62	44-78	≥80	≤42
64	46-80	≥82	≤44
66	48-82	≥84	≤46
68	50-84	≥86	≤48
70	52-86	≥88	≤50
72	54-86	≥88	≤52
74	56-88	≥90	≤54
76	58-90	≥92	≤56
78	60-92	≥94	≤58
80	64-92	≥94	≤62
82	66-94	≥96	≤64
84	68-94	≥96	≤66
86	70-96	≥98	≤68

Continued

was 16% in patients with baseline WRS ≥ 90%.[27] For trials using hearing response as the primary outcome, the functional group recommends excluding patients with WRS ≥ 90% to allow for appreciation of hearing improvement and to ensure that test-to-test variation does not obscure true changes in hearing function.

Frequency of reevaluation. No evidence-based guidelines are available to help determine the interval between evaluations. The functional group recommends reevaluation of maximum WRS every 3 months during initial treatment and every 3–6 months during prolonged treatment.

Confirmatory measurement/duration of response. In nonrandomized trials where response is the primary endpoint, confirmation of response is desirable. The time between audiograms should be at least 1 month. Responses that are not confirmed by subsequent audiology should be termed "unconfirmed response." For randomized trials or for trials where time to hearing decline is the primary endpoint, confirmation of response is not required. To declare stable disease, audiograms must show stable disease at least once after study initiation for a period of time that is defined by the study protocol (usually 6 months).

Duration of hearing response. The duration of HR is measured from the time when study measurement first shows a HR until the first date that maximum WRS is recorded within the 95% critical difference of the baseline WRS.

Duration of stable hearing. SH is measured from the start of treatment until the criteria for hearing decline are met, taking as reference the 95% critical difference of the baseline maximum WRS. For protocols designed to study the proportion of patients with SH for a minimum period of time, the protocol should specify the interval between audiograms.

It is important to note that the duration of HR/SH and the determination of time to hearing decline is influenced by the frequency of follow-up after baseline evaluation. These intervals should be defined in the study protocol.

Proportion free from hearing loss. Historically, most patients with NF2 experience profound bilateral hearing loss during the course of life. For young patients and those with good hearing, a therapy that could prevent hearing loss would be desirable, even if this therapy did not improve hearing. In this scenario, the proportion of patients free from hearing loss at landmark time points might be considered appropriate to provide an initial sign of biologic activity. However, in uncontrolled phase II trials this endpoint is subject to bias from known and unknown factors that influence hearing loss. Ideally, this endpoint is best evaluated in the setting of a randomized trial. In the case of NF2, spontaneous hearing improvement is sufficiently uncommon (about 16% for patients with

Table 2	Continued		
Baseline word recognition score (%)	95% critical difference (%)	Hearing response (%)	Progressive hearing loss (%)
88	74–96	≥98	≤72
90	76–98	100	≤74
92	78–98	100	≤76
94	82–98	100	≤80
96	86–100	n/a	≤84
98	90–100	n/a	≤88
100	96–100	n/a	≤94

Abbreviation: n/a = not applicable.
Clinical criteria for definition of hearing response based on a 50-item monosyllable word recognition test. Upper and lower limits for the 95% critical differences for percentage scores are adapted from Thornton.[35]

baseline WRS < 90%) that a nonrandomized trial may be justifiable.[27]

Reporting best response results. For phase II trials, all patients in the study must be included in the report of the results, even those who are not evaluable or experience major protocol deviations. Patients should be designated as either HR, SH, hearing decline, or not evaluable (e.g., early death, toxicity, assessment not performed). In general, the denominator for calculation of the response rate should include all eligible patients. Conclusions should not be based on selected "evaluable" patients. For phase III trials, hearing response evaluation may be a primary or secondary endpoint. If hearing response evaluation is a primary endpoint (i.e., the study includes only patients who are capable of hearing improvement), the discussion above should apply. If response evaluation is a secondary endpoint (i.e., the study includes all patients regardless of hearing), the hearing response rate may be reported using an "intention-to-treat" analysis (with all randomized patients in the denominator) or using a predefined subset analysis (including the subset of patients with hearing loss at baseline). The study protocol should specify how response results will be reported, including any planned subset analyses.

Other outcomes for hearing: Pure-tone average. An alternative measure of hearing includes simple detection of sound (i.e., the threshold of audibility). Pure-tone thresholds are most often measured at the frequencies of 250, 500, 1,000, 2,000, 3,000, 4,000, 6,000, and 8,000 Hz. The PTA is calculated as the average of thresholds at 500, 1,000, 2,000, and 4,000 Hz,[8] although other thresholds may be used.[28] Changes in pure-tone audiometry have been used in oncology studies to monitor ototoxicity from chemotherapy. However, in patients with VS, word recognition is affected to a greater extent than detection of pure tones[29] due to cochlear and neural alterations.[30] Furthermore, patients' quality of life is directly limited by reduced word recognition rather than audibility, which can be addressed by amplification. For this reason, the committee does not endorse changes in PTA as a primary outcome for hearing trials but does recommend that PTA be recorded as a secondary outcome.

Recommended primary outcome for facial function: SMILE analysis. SMILE analysis is a technique to measure facial function that relies on computerized measurement of facial excursion based on photographs taken in standard positions (figure). The system uses the iris diameter (corneal white-to-white diameter) as a scale reference for all facial dimensions on the same photograph. This built-in scale—in the same plane as the oral commissure—allows measurements of horizontal to vertical commissure excursion to be extrapolated from frontal photographs using readily available photo-editing software.
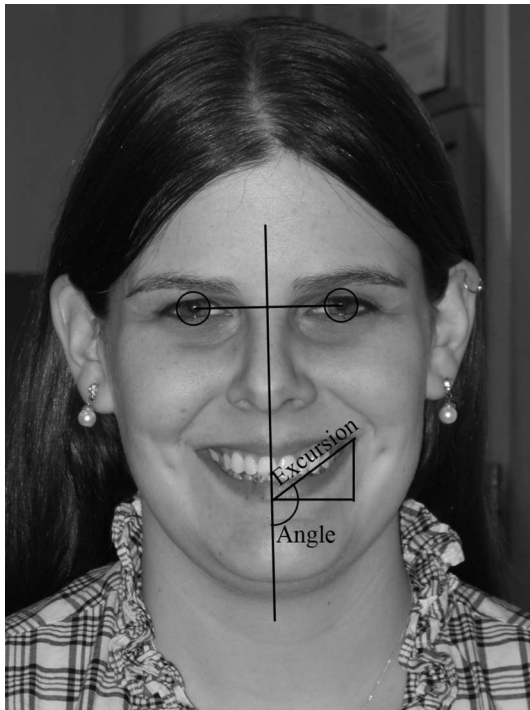
Feasibility. SMILE analysis is performed using frontal photographs of the patient at rest and with a full smile. Photos are imported into the software program (Facegram), which then calculates an objective analysis of smile symmetry and change in commissure excursion. The Java-based SMILE analysis program is free for use and can be downloaded at www.sircharlesbell.org/facial_nerve_programs.html. Complete facial analysis takes less than 5 minutes and may be stored electronically for future use.

Patient characteristics and validity. SMILE analysis can be reliably performed in individuals capable of standard facial poses.[17] The lower age limit for reliability of measurement has not been determined. In a previous study of 10 normal individuals (20 hemi-smiles), there was a strong correlation between measured distances and the true excursion (class correlation coefficients $R = 0.96$–0.99). There was strong correlation between the measurements of excursion distances for multiple raters (intraclass correlation coefficients: $R = 0.88$–0.98). The intratest reliability after 1 week (for a given rater) was 0.99, indicating a strong correlation between measurements over time.

Use in published studies. SMILE analysis has been used as an endpoint in studies of facial reanimation due to many causes, including NF2.[31,32] To date, the technique has not been used in multicenter studies.

Baseline documentation of "target" and "nontarget" hemi-smiles. When unilateral facial weakness is present at baseline, the target hemi-smile should be identified and recorded at baseline. In this instance, the contralateral (nontarget) hemi-smile acts as a normal control for the affected side. A history of previous surgery that might affect facial strength (e.g., resection of VS)

such as keratopathy. SMILE analysis of facial photos is capable of measuring corneal exposure before and after intervention. Measuring changes in lagophthalmos should be studied with a goal of defining response criteria in the future. Presently, the group recommends a 2-mm increase in lid position as a functional response and a 2-mm decrease in lid position as functional decline.

Frequency of reevaluation. No evidence-based guidelines are available to determine an appropriate interval between evaluations. The functional group recommends reevaluation of true smile excursion every 3 months during the first year after surgical intervention and yearly afterwards.

Confirmatory measurement/duration of response. In nonrandomized trials where response is the primary endpoint, confirmation of response is desirable. The minimum time between facial analyses should be at least 1 month. Responses not confirmed by subsequent facial photos are termed "unconfirmed response." For randomized trials or for trials where time to facial weakness is the primary endpoint, confirmation of response is not required. To declare stable facial strength, facial photos must show stable strength at least once after intervention at a time that is defined by the study protocol.

Duration of facial response. The duration of facial response is measured from the time when study measurement first shows a facial response until the first date that facial photos demonstrate progressive facial weakness compared to the best change in true smile excursion.

Duration of stable facial function. Stable facial function is measured from the start of treatment or intervention until the criteria for progressive facial weakness are met, taking as reference the true smile excursion at baseline. For protocols designed to study the proportion of patients with stable facial function for a minimum period of time, the protocol should specify the interval between facial photos.

The duration of smiling response/stable facial function and the determination of time to progressive facial weakness are influenced by the frequency of follow-up after baseline evaluation. These intervals should be defined in the study protocol.

Proportion free from progressive facial weakness. Historically, patients with NF2 experience some facial weakness over their lifetime. For patients with normal facial function, a novel procedure or technique (e.g., surgery or radiation) that minimizes, delays, or otherwise prevents facial weakness would be desirable. In this scenario, the proportion of patients free from progressive facial weakness at landmark time points could be an appropriate measure of successful intervention. However, in uncontrolled phase II trials this endpoint is subject to bias from known and

should be recorded. When bifacial weakness is present at baseline, a target hemi-smile should be identified and recorded at baseline, based on normative smile parameters.[33]

Patients should be photographed using at least 2 facial poses, including "at rest" and with a broad smile. Using the software, the true smile excursion from midline and the smile angle from horizontal midline should be recorded for target and nontarget hemi-smiles (figure). Comparison between smile angles and smile excursion on the normal side and the affected side at baseline provides an objective analysis of smile symmetry. Comparison between smile angles and smile excursion of the target hemi-smile before and after intervention provides an objective analysis of change in commissure excursion with treatment.

Response criteria. Consensus criteria using SMILE analysis (or other facial rating scales) have not been defined to date. Until more data are acquired, the functional committee proposes to define a response as an increase in true smile excursion of 25% compared with the true smile excursion at baseline. Progression is defined as a decrease in true smile excursion of 25% compared with the baseline true smile excursion. Stable function is defined as all other changes in true smile excursion compared with baseline. These response criteria apply to patients with both unilateral facial weakness and bifacial weakness.

Special notes on assessment of facial function. Patients with facial weakness also experience paralytic lagophthalmos, which can increase risk of complications

unknown factors that influence facial function. Ideally, this endpoint is best evaluated in the setting of a randomized trial.

Reporting best response results. For phase II trials, all patients in the study must be included in results reporting, even those who are not evaluable or experience major protocol deviations. Patients should be designated as either facial response, stable facial function, progressive facial weakness, or not evaluable (e.g., early death, toxicity, assessment not performed). In general, the denominator for calculation of the response rate should include all eligible patients. Conclusions should not be based on selected "evaluable" patients. For phase III trials, response evaluation may be a primary or secondary endpoint. If response evaluation is a primary endpoint (i.e., the study includes only patients who have baseline facial weakness), the discussion above should apply. If response evaluation is a secondary endpoint (i.e., the study includes all patients regardless of facial function), the response rate may be reported using an "intention-to-treat" analysis (with all randomized patients in the denominator) or using a predefined subset analysis (including the subset of patients with facial weakness at baseline). The study protocol should specify how response results will be reported, including any planned subset analyses.

Other outcomes for facial function: House-Brackmann scale. As noted above, facial function in patients with VS is commonly monitored using the HB scale.[10] Given the widespread use of this scale and its correlation with other scales of facial function,[34] the committee recommends that the HB score be recorded as a secondary outcome for trials of NF2 patients with VS.

CONCLUSION These recommendations by the REiNS Collaboration are intended to provide researchers with a common set of endpoints for clinical trials for patients with NF. The use of shared endpoints will improve the quality of clinical trials and encourage pharmaceutical companies to invest in drug development in this patient population. The REiNS International Collaboration recognizes that no criteria are perfect and that alternative endpoints can be justified for these studies. In the future, the group intends to update these recommendations as more information is published about clinically relevant endpoints.

AUTHOR CONTRIBUTIONS

SR Plotkin: drafting the manuscript, study concept, interpretation of data. SL Ardern-Holmes: revising the manuscript for content, study concept, interpretation of data. FG Barker: drafting the manuscript, study concept, interpretation of data. JO Blakeley: revising the manuscript for content, study concept, interpretation of data. DG Evans: revising the manuscript for content, study concept, interpretation of data. RE Ferner: revising the manuscript for content, study concept, interpretation of data. TA Hadlock: revising the manuscript for content, study concept, interpretation of data. C Halpin: drafting the manuscript, study concept, interpretation of data.

DISCLOSURE

REFERENCES

1. Evans DG. Neurofibromatosis 2. In: Pagon RA, Bird TD, Dolan CR, editors. Seattle: University of Washington; 2011.
2. Neary WJ, Hillier VF, Flute T, Stephens D, Ramsden RT, Evans DG. Use of a closed set questionnaire to measure primary and secondary effects of neurofibromatosis type 2. J Laryngol Otol 2010;124:720–728.
3. Neary WJ, Hillier VF, Flute T, Stephens SD, Ramsden RT, Evans DG. The relationship between patients' perception of the effects of neurofibromatosis type 2 and the domains of the Short Form-36. Clin Otolaryngol 2010;35:291–299.

4. Hornigold R, Golding J, Leschziner G, et al. The NFTI-QOL: a disease-specific quality of life questionnaire for neurofibromatosis 2. J Neurol Surg B Skull Base 2012; 73:104–111.

5. Masuda A, Fisher LM, Oppenheimer ML, Iqbal Z, Slattery WH. Hearing changes after diagnosis in neurofibromatosis type 2. Otol Neurotol 2004;25:150–154.

6. Tysome JR, MacFarlane R, Durie-Gair J, et al. Surgical management of vestibular schwannomas and hearing rehabilitation in neurofibromatosis type 2. Otol Neurotol 2012;33:466–472.

7. Friedman RA, Goddard JC, Wilkinson EP, et al. Hearing preservation with the middle cranial fossa approach for neurofibromatosis type 2. Otol Neurotol 2011;32:1530–1537.

8. Halpin C. Measuring audiometric outcomes. In: Shin J, Hartnick C, Randolph G, editors. Evidence Based Otolaryngology. New York: Springer; 2008:227–238.

9. Bess FH, Josey AF, Humes LE. Performance intensity functions in cochlear and eighth nerve disorders. Am J Otol 1979;1:27–31.

10. House JW, Brackmann DE. Facial nerve grading system. Otolaryngol Head Neck Surg 1985;93:146–147.

11. Vrabec JT, Backous DD, Djalilian HR, et al. Facial nerve grading system 2.0. Otolaryngol Head Neck Surg 2009; 140:445–450.

12. Kang TS, Vrabec JT, Giddings N, Terris DJ. Facial nerve grading systems (1985-2002): beyond the House-Brackmann scale. Otol Neurotol 2002;23:767–771.

13. American Academy of Otolaryngology-Head and Neck Surgery Foundation I. Committee on hearing and equilibrium guidelines for the evaluation of hearing preservation in acoustic neuroma (vestibular schwannoma). Otolaryngol Head Neck Surg 1995;113:179–180.

14. Gardner G, Robertson JH. Hearing preservation in unilateral acoustic neuroma surgery. Ann Otol Rhinol Laryngol 1988;97:55–66.

15. Halpin C, Rauch SD. Using audiometric thresholds and word recognition in a treatment study. Otol Neurotol 2006;27:110–116.

16. Brackmann DE, Owens RM, Friedman RA, et al. Prognostic factors for hearing preservation in vestibular schwannoma surgery. Am J Otol 2000;21:417–424.

17. Bray D, Henstrom DK, Cheney ML, Hadlock TA. Assessing outcomes in facial reanimation: evaluation and validation of the SMILE system for measuring lip excursion during smiling. Arch Facial Plast Surg 2010;12:352–354.

18. Lee LN, Susarla SM, Hohman H, Henstrom DK, Cheney ML, Hadlock TA. A comparison of facial nerve grading systems. Ann Plast Surg 2013;70:313–316.

19. Bess F, Humes L. Audiology: The Fundamentals, 4th ed. Philadelphia: Lippincott Williams and Wilkins; 2008.

20. Halpin C. Audiologic evaluation and hearing ids. In: Lazar R, editor. Neurorehabilitation. New York: McGraw-Hill; 1998:649–676.

21. De Cardenas M, Marrero V. Cuaderno de logoaudiometria. Madrid: Universidad Nacional de Educacion a Distancia; 1994.

22. Plotkin SR, Stemmer-Rachamimov AO, Barker FG, et al. Hearing improvement after bevacizumab in patients with neurofibromatosis type 2. N Engl J Med 2009;361:358–367.

23. Plotkin SR, Merker VL, Halpin C, et al. Bevacizumab for progressive vestibular schwannoma in neurofibromatosis type 2: a retrospective review of 31 patients. Otol Neurotol 2012;33:1046–1052.

24. Karajannis MA, Legault G, Hagiwara M, et al. Phase II trial of lapatinib in adult and pediatric patients with neurofibromatosis type 2 and progressive vestibular schwannomas. Neuro Oncol 2012;14:1163–1170.

25. Hirsh IJ, Davis H, Silverman SR, Reynolds EG, Eldert E, Benson RW. Development of materials for speech audiometry. J Speech Hear Disord 1952;17:321–337.

26. Halpin C, Thornton AR. Q/MASS Volume 2: recording of speech audiometry materials. Minneapolis: Massachusetts Eye and Ear Infirmary/Qualitone Inc; 1991.

27. Plotkin SR, Merker VL, Muzikansky A, Barker FG II, Slattery W III. Natural history of vestibular schwannoma growth and hearing decline in newly diagnosed neurofibromatosis type 2 patients. Otol Neurotol 2013. In press.

28. Monsell EM. New and revised reporting guidelines from the Committee on Hearing and Equilibrium. American Academy of Otolaryngology-Head and Neck Surgery Foundation, Inc. Otolaryngol Head Neck Surg 1995; 113:176–178.

29. Caye-Thomasen P, Dethloff T, Hansen S, Stangerup SE, Thomsen J. Hearing in patients with intracanalicular vestibular schwannomas. Audiol Neurootol 2007;12:1–12.

30. Mahmud MR, Khan AM, Nadol JB Jr. Histopathology of the inner ear in unoperated acoustic neuroma. Ann Otol Rhinol Laryngol 2003;112:979–986.

31. Vakharia KT, Henstrom D, Plotkin SR, Cheney M, Hadlock TA. Facial reanimation of patients with neurofibromatosis type 2. Neurosurgery 2012;70:ons237–ons243.

32. Hadlock TA, Malo JS, Cheney ML, Henstrom DK. Free gracilis transfer for smile in children: the Massachusetts eye and ear infirmary experience in excursion and quality-of-life changes. Arch Facial Plast Surg 2011; 13:190–194.

33. Hadlock TA, Urban LS. Toward a universal, automated facial measurement tool in facial reanimation. Arch Facial Plast Surg 2012;14:277–282.

34. Berg T, Jonsson L, Engstrom M. Agreement between the Sunnybrook, House-Brackmann, and Yanagihara facial nerve grading systems in Bell's palsy. Otol Neurotol 2004;25:1020–1026.

35. Thornton AR, Raffin MJ. Speech-discrimination scores modeled as a binomial variable. J Speech Hear Res 1978;21:507–518.

Neurology®

Hearing and facial function outcomes for neurofibromatosis 2 clinical trials

Scott R. Plotkin, Simone L. Ardern-Holmes, Fred G. Barker II, et al.

This information is current as of November 18, 2013

AMERICAN ACADEMY OF NEUROLOGY®