Comparison between Subjective Scoring and Computer-Based Asymmetry Assessment in Facial Nerve Palsy

Doh Young Lee[1]*, Hyun Seok Kim[2]*, So Young Kim[3], Kwang Suk Park[4], and Young Ho Kim[1]

[1]Department of Otorhinolaryngology Head and Neck Surgery, Seoul National University Boramae Medical Center, Seoul, Korea
[2]Interdisciplinary Program of Bioengineering, Seoul National University, Seoul, Korea
[3]Department of Otorhinolaryngology Head and Neck Surgery, CHA Bundang Medical Center, CHA University, Seongnam, Korea
[4]Department of Biomedical Engineering, College of Medicine, Seoul National University, Seoul, Korea

Background and Objectives: The aim of the present study was to assess the feasibility of a PC-based facial asymmetry assessment program (PC-FAAP) and to compare the results of PC-FAAP with subjective regional scoring by raters in acute unilateral peripheral facial nerve paralysis (FNP). **Subjects and Methods**: Participants were divided into 3 groups with 8 participants per group: group I, normal; group II, mild to moderate FNP; and group III, severe FNP. Using the PC-FAAP, the mouth asymmetry ratio (MAR), eyebrow asymmetry ratio (EAR), and complete eye closure asymmetry ratio (CAR) were calculated by comparing the movement of tracking points on both sides. The FNP grading scale (FGS) integrated each score, and the scores were weighted with a ratio of 5:3:2 (MAR:CAR:EAR). Subjective regional scoring was measured on a 0−100 scale score by three otologists. PC-FAAP and subjective scoring were compared in each group regarding the consistency of the results. **Results**: The mean scores of the MAR, EAR, CAR, and FGS of each group were significantly different. PC-FAAP showed significant differences between the three groups in terms of MAR, EAC, CAR, and FGS. PC-FAAP showed more consistent results than subjective assessment (*p*<0.001). The PC-FAAP was significantly more consistent in group I and group III (*p*<0.001 and *p*=0.002, respectively). FGS in group III was the only parameter that showed a more consistent result in PC-FAAP than the subjective scoring (*p*=0.008). **Conclusions**: An FNP grading system using a PC-based program may provide more consistent results, especially for severe forms. **J Audiol Otol 2019;23(1):53-58**

KEY WORDS: Facial nerve palsy · Assessment · Asymmetry.

Introduction

Facial nerve paralysis (FNP) appears as a weakness or loss of facial muscle expression due to damage of the facial nerve. FNP with incomplete recovery significantly influences patients' quality of life. Facial asymmetry may induce a disturbance of social relationships due to a cosmetic problem and inappropriate eye care or perioral muscular recovery may present functional disturbances. FNP patients require support from many medical specialists, such as neurologists, otolaryngologists, ophthalmologists, plastic surgeons, psychiatrists, and physio-therapists, to appropriately manage FNP [1]. Although patients with acute peripheral FNP (such as Bell's palsy) achieve a high rate of complete remission, some fail to achieve recovery and suffer from sequelae after FNP. Early diagnosis and proper grading are crucial in determining the prognosis of FNP and in optimally managing the disease.

Various grading systems have been introduced to assess degrees of FNP. The House-Brackmann grading system and Fisch scaling are examples of the most widely used tools. However, an objective, quantitative, and integrated FNP assessment tool has not been established to date. The complexity of assessment, lack of objectivity and reproducibility, test-retest reliability, and inter-observer differences were suggested as problems of these grading systems. Studies have explicitly attempted to improve the reproducibility, result consistency, and ease of

performance of FNP assessment tools [2,3]. The objective and consistent assessment of FNP requires minimizing inter-observer variation.

The present study compared the results of a PC-based facial asymmetry assessment program (PC-FAAP) developed to objectively calculate and standardize FNP grading with subjective regional scoring by raters in three facial expression components (smiling, forced closing eyes, and raising eyebrows) for patients with acute unilateral peripheral FNP.

Subjects and Methods

This study was approved by the Institutional Review Board (IRB) of the Seoul National University Boramae Medical Center (IRB No. 16-2016-83). Informed consent was obtained from all participants.

Facial photographs

The facial photographs of the participants were acquired using a camera with a 1334×750 resolution. The frontal photographs of the participants standing at a consistent distance were taken using four facial movements as FNP static views: "At rest," "Large smile showing teeth," "Closure of eyes tightly," and "Elevation of eyebrows" [4-6].

PC-FAAP

Facial photographs (resting, mouth extension, forced eye closing, and eyebrow lifting) of 8 healthy people and 16 FNP patients were used. Participants were divided into 3 groups in accordance with the House-Brackmann grading (H-B): group I (normal)=H-B grade I−II, group II (mild to moderate FNP)=H-B grade III−IV, and group III (severe FNP)=H-B grade V−VI. The facial asymmetry was evaluated using PC-FAAP, which we previously developed. This program determined primary facial landmarks and tracked them with an incremental parallel cascade of the linear regression method proposed by Asthana, et al. [7]. The localization algorithms of facial landmarks, data acquisition, and analysis methods were determined as described in our previous study [8]. We defined horizontal and vertical lines to assess facial asymmetry. The horizontal line was defined as an extended line connecting each medial canthus. The vertical line was defined as a line perpendicular to the horizontal line at a mid-point between both medial canthus points.

The mouth asymmetry ratio (MAR) was calculated by comparing both sides of the tracking pointed lip area centered around the vertical midline (Fig. 1, upper). The eyebrow asymmetry ratio (EAR) was measured using the vertical distance from the midpoint of the medial and lateral canthal line to the up-

per margin of the eyebrow when lifting the eyebrows (Fig. 1, middle). The complete eye closure asymmetry ratio (CAR) was estimated using the area between a point-tracked low eyeline and the upper margin of the eyebrow with an identical and parallel distance from each medial canthal point (Fig. 1, lower). A denominator of each ratio in a normal healthy group was determined as the larger of the values on both sides. Our previous study revealed that the perception of the most unnat-
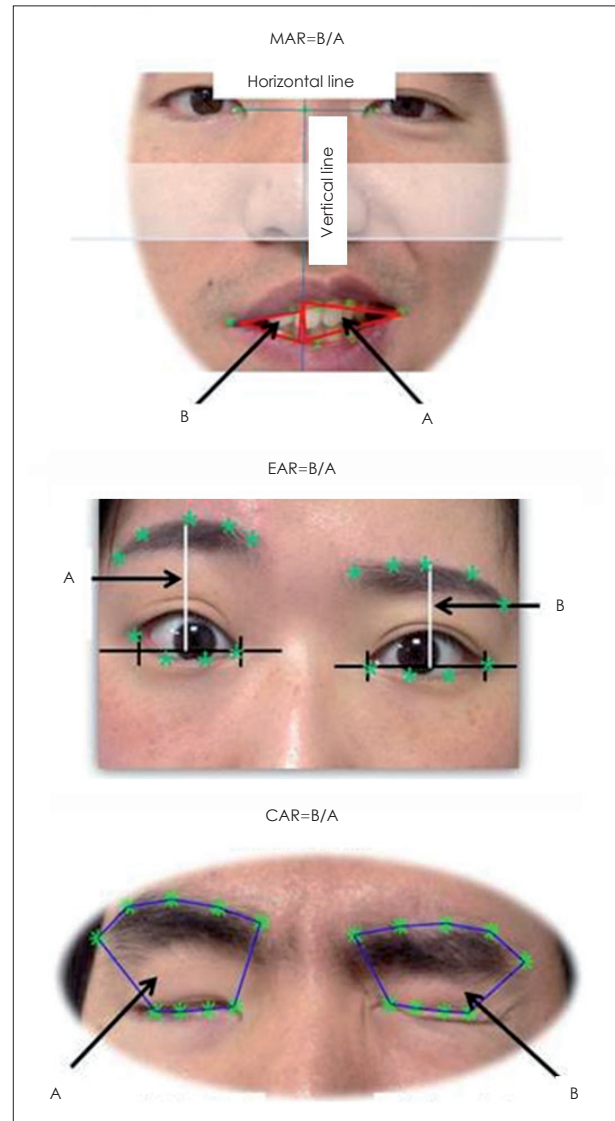

Fig. 1. Summary of PC-based facial asymmetry assessment program scaling methods. The MAR was calculated by the relative proportion of the opened mouth area on the paralyzed side to that on the non-paralyzed side (upper). The EAR was calculated by the relative proportion of movement distance of the eyebrow on the paralyzed side to that on the non-paralyzed side (middle). CAR was calculated by the relative proportion of the area between the point-tracked low eye line and the upper margin of the eyebrow on the paralytic side to that of the non-paralyzed side (lower). MAR: mouth asymmetry ratio, EAR: eyebrow asymmetry ratio, CAR: complete eye closure asymmetry ratio.

Table 1. Mean score of MAR, EAR, CAR, and FGS

	Group I	Group II	Group III	Group I vs. group II/III	Group II vs. group III
MAR	0.93 (0.82−0.99)	0.77 (0.55−0.98)	0.39 (0.18−0.67)	$p<0.001$	$p=0.001$
EAR	0.97 (0.94−1.00)	0.87 (0.81−0.95)	0.74 (0.64−0.82)	$p<0.001$	$p<0.001$
CAR	0.96 (0.89−0.99)	0.87 (0.80−0.93)	0.77 (0.64−0.85)	$p<0.001$	$p=0.007$
FGS	0.95 (0.89−0.98)	0.82 (0.70−0.94)	0.67 (0.43−0.75)	$p<0.001$	$p=0.001$

MAR: mouth asymmetry ratio, EAR: eyebrow asymmetry ratio, CAR: complete eye closure asymmetry ratio, FGS: facial nerve paralysis grading score

ural facial area in FNP patients was the mouth, followed by the eyes and the eyebrows [9]. Accordingly, MAR, CAR, and EAR were weighted with a ratio of 5:3:2 to determine the FNP grading score (FGS) while considering the results of perception of facial asymmetry according to the facial area found in our previous study [9].

We compared differentiation in facial asymmetry using this program for three groups: normal, mild to moderate, and severe FNP. To evaluate the consistency of the results, the three scores of each subject consecutively measured by PC-FAAP were compared to the visual analogue scale (VAS) scores measured by three otologists who were not aware of this study. They were surveyed to rate the grade of FNP compared to the unaffected side on a 100-scale score for mouth extension, forced eye closing, and eyebrow lifting. VAS 100-score indicated that there is no asymmetry. Additionally, 100-scaled scores for overall FNP grade compared to the unaffected side were also presented.

Statistical analysis

The MAR, EAR, CAR, and FGS were compared between groups using the Mann-Whitney U test. The correlations between resting and expressed MAR or EAR were investigated using Pearson correlation analyses. A variation analysis was performed using one-way ANOVA. The SPSS 21.0 statistical package was used for the analysis (IBM Corp., Armonk, NY, USA). p values ≤0.05 were considered statistically significant.

Results

Validation of PC-FAAP

The mean scores of MAR, EAR, CAR, and FGS are shown in Table 1, and scores between groups showed significant differences. In all parameters, statistical significance was attained when comparing groups I and II/III ($p<0.001$) and between groups II and III ($p=0.001$, $p<0.001$, $p=0.007$, and $p=0.001$ for MAR, EAR, CAR, and FGS, respectively).

Consistency assessment of PC-FAAP

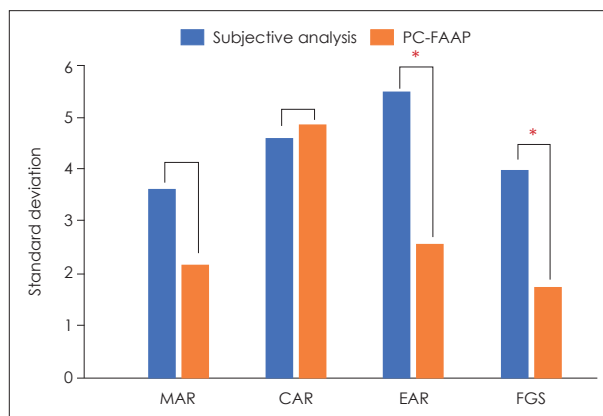MAR was more consistent in PC-FAAP; however, it failed to achieve statistical significance ($p=0.085$). EAR and FGS

Fig. 2. Comparison of standard deviation in the total cohort and each grading score (MAR, CAR, EAR, and FGS). PC-FAAP was more consistent in the total cohort and EAR and FGS scores. *$p<0.05$. PC-FAAP: PC-based facial asymmetry assessment program, MAR: mouth asymmetry ratio, EAR: eyebrow asymmetry ratio, CAR: complete eye closure asymmetry ratio, FGS: facial nerve paralysis grading score.
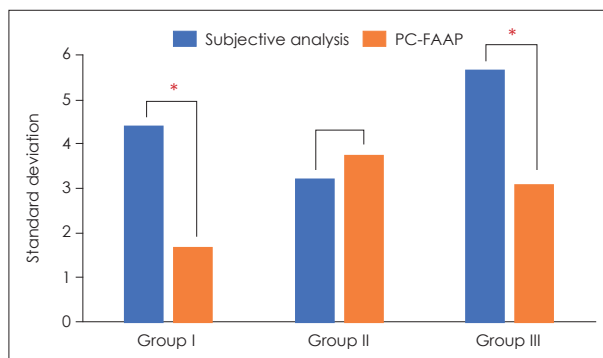

Fig. 3. Comparison of standard deviation in each group. Groups I and III showed more consistency in PC-FAAP than subjective analysis. *$p<0.05$. PC-FAAP: PC-based facial asymmetry assessment program.

showed more significant measurement consistency in PC-FAAP than subjective assessment ($p=0.004$ and $p<0.001$, respectively) (Fig. 2). Regarding comparisons between each group, PC-FAAP demonstrated significantly more consistent results than subjective assessment in groups I and III ($p<0.001$ and $p=0.002$, respectively) (Fig. 3). The assessment of FNP using PC-FAAP showed better consistency in FGS than subjective assessment in group III ($p=0.008$) (Fig. 4).
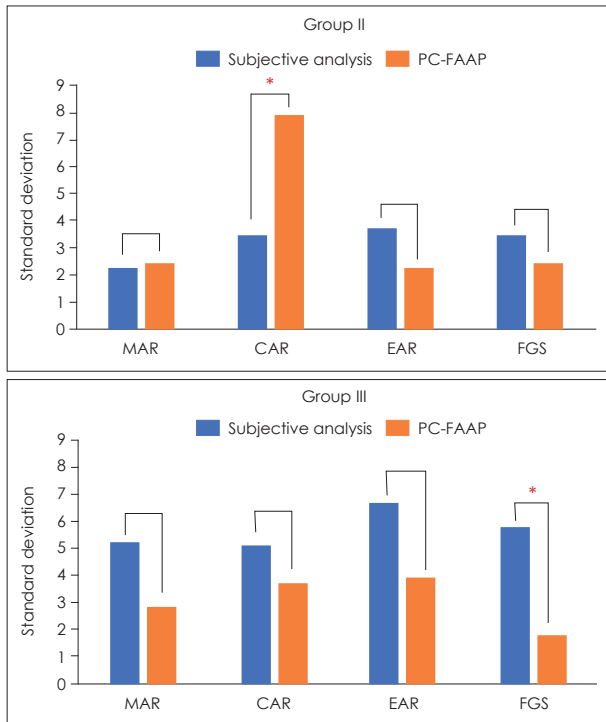
Fig. 4. Comparison of standard deviation for each grading score in patients with facial nerve paralysis (groups II and III). FGS in group III was the only parameter that showed significantly better consistency than subjective analysis. *p<0.05. PC-FAAP: PC-based facial asymmetry assessment program, MAR: mouth asymmetry ratio, EAR: eyebrow asymmetry ratio, CAR: complete eye closure asymmetry ratio, FGS: facial nerve paralysis grading score.

Discussion

Botman and Jongkees [10] introduced the first FNP scaling system with 5 grades in 1955. In 1983, House-Brackmann [11] developed a new scale to identify the status and recovery of facial nerve damage after removing vestibular schwannomas, which has been a widely used scale of FNP to date. However, this well-made scaling system is being used in various diseases showing FNP such as Bell's palsy, Ramsay-Hunt syndrome, and FNP by infection, trauma, or tumor. There are also at least 10 or more scaling systems [11,12]. Most scaling systems have some drawbacks. Among others, the complexity of assessment methods and non-objectivity of rater measurements appear to be the most severe weakness for existing FGS systems. A lack of objectivity in grading FNP may complicate comparing functional changes and sharing accurate information. Measuring and scaling FNP by an automatized program, not by raters, can reduce the individual difference that raters make.

We propose an FGS for patients with FNP using a PC-based program in three facial expression components: large smile showing teeth, closure of eyes tightly, and elevation of the eyebrows, without constraints from the recording environ-ment. Although the survey study reported "Closure of eyes gently" instead of "Closure of eyes tightly" as a view described in common grading scales, we chose to use "Closure of eyes tightly," which can better reflect the function of the orbicular-is oculi [4,6]. Although the present study did not include var-ious variables such as static, dynamic, synkinesis, and facial area, this preliminary scaling could contribute to early diag-nosis and treatment in patients with FNP when considering the current feasibility of wireless communication. In addition, this integrated assessment of facial asymmetry at facial expres-sion points may be helpful in evaluating the severity of FNP and improving existing FNP grading systems.

The PC-based program used in the present study was de-veloped to assess FNP patients using clear and straightfor-ward methods and to avoid interrater unreliability. It is rela-tively easy to design and modify modules using this program. Additionally, because facial photographs of patients with FNP are entered into the program and analyzed, we could plan to establish remote medical examination and treatment in fur-ther research. A PC-based program can be used by non-prac-titioners to enable early FNP detection and diagnosis. Our study demonstrated that severe FNP patients were the best candidates for PC-based analysis. Although participants with normal facial expressions also showed more consistent results with PC-based analysis, we think that the system is not useful for treating FNP in this context; however, it may be meaning-ful if used in aesthetic procedures.

Although there was a small number of participants in each group, two significant results should be emphasized. First, the PC-based assessment was significantly consistent in measur-ing eyebrow elevation. As shown in Fig. 2, the MAR was also more consistent in PC-FAAP; however, we think that statisti-cal significance was not attained due to the small number of patients. Notably, the maximal eye closure assessment was ap-proximately the same between PC-FAAP and the subjective analysis in terms of consistency. Consistency in the subjec-tive analysis of facial expression has been reported to be best for eye closure [13]. Therefore, we could conclude that PC-based assessment is more advantageous in assessing eye-brow or mouth movement and that subjective assessment can be as consistent as a PC-based assessment. However, there were no significant differences in the consistency of the sub-group analysis. Considering the small number of patients in-cluded, a larger cohort study would reveal significant differ-ences between groups II and III.

Second, the PC-based and subjective assessments were not significantly different in mild to moderate FNP patients with regard to consistency, while a PC-based assessment was su-perior in severe FNP patients (Fig. 3 and 4). This was con-

trary to our expectations when we started this study. We hypothesized that PC-FAAP might help consistently identify and differentiate mild FNP; however, our study showed that PC-FAAP had comparable variation in assessing mild to moderate FNP. Moreover, the subjective assessment was more consistent in CAR, and the superiority of the subjective assessment found in evaluating eye closure can be a possible reason. Although the exact reason for this finding was not be discussed further, PC-FAAP is more appropriate and advantageous in measuring facial expressions in severe FNP cases, and further studies are needed to improve the consistency of PC-FAAP in mild to moderate FNP patients. Statistical significance was not attained in severe FNP (group III); however, MAR, CAR, and EAR showed better consistency with PC-FAAP. We think that a study with a larger cohort may show statistical significance.

Elderly subjects with severe FNP may have a substantial confounding factor when evaluating facial nerve function by pixel subtraction using a PC program [14]. The authors suggested that their skin laxity and facial hypomobility can induce errors in this system. In addition, the weighting of facial grading variables may be arbitrary, although the ratio used in our study is based on our previous study [9]. People reported that the most unnatural facial areas (unnatural facial expression) in FNP patients were the lip (smile) followed by the eye (eye closure), cheek (whistling), and eyebrow (frowning). However, this is based on subjective assessment, and weighting variables can be changed by racial, ethnic, or cultural differences. Further studies regarding weighting on each area of the paralyzed face will indicate more representative scores. Facial grading variables, such as MAR and EAR at rest status, were not significantly different between groups. This may mean that PC-FAAP is not sensitive to FNP at rest, which should be refined in the future.

The present study has some limitations. 1) Facial asymmetry evaluation in FNP patients was conducted in a single center. Further study in multiple centers may redeem the variations generated by the calibration. 2) This study was limited to the static measurement of faces with FNP. Further developing PC-FAAP using facial videos may improve the accuracy of FNP grading variables. 3) Because PC-FAAP is a tentative program, a supplemental point tracking technique is needed for the accurate detection of facial asymmetry. 4) PC-FAAP did not analyze other considerations, including synkinesis, in this study. Further studies should apply these to obtain more detailed results. 5) Bilateral FNP cannot be evaluated with this program, as is the case with most other assessment tools. This may be another problem because the present grading systems depend on comparisons with the non-affected side [15].

The following are required to commoditize the FNP assessment tool: 1) Simplification of the assessing method, 2) Qualification and objectification of scaling, and 3) Establishment of consensus for the sharing of practitioners. To fulfill this, further studies that measure and analyze various conditions of the FNP should be performed. Additionally, the authorized organization should conduct verification research to assess programs that show effectiveness.

In conclusion, the PC-based assessment of FNP can avoid objective scoring data bias by raters. This assessment may be more advantageous than eyebrow asymmetry and integrated FGS, and the grade of severely affected FNP can be assessed more consistently by PC-FAAP. An objective grading system using PC may be more appropriate in severe FNP patients than mild to moderate FNP patients. Additional studies with more patients may further reinforce the functions of PC-FAAP.

Conflicts of interest

The authors have no financial conflicts of interest.

REFERENCES

1) Eviston TJ, Croxson GR, Kennedy PG, Hadlock T, Krishnan AV. Bell's palsy: aetiology, clinical features and multidisciplinary care. J Neurol Neurosurg Psychiatry 2015;86:1356-61.

2) Chong LSH, Eviston TJ, Low TH, Hasmat S, Coulson SE, Clark JR. Validation of the clinician-graded electronic facial paralysis assessment. Plast Reconstr Surg 2017;140:159-67.

3) Zago M, Tarabbia F, Bassetti L, Biglioli F, Sforza C. Photographic technique for the quantitative assessment of lagophthalmos and eyelid position in patients with facial palsy. Br J Oral Maxillofac Surg 2017;55:548-50.

4) Santosa KB, Fattah A, Gavilán J, Hadlock TA, Snyder-Warwick AK. Photographic standards for patients with facial palsy and recommendations by members of the Sir Charles Bell Society. JAMA Facial Plast Surg 2017;19:275-81.

5) Barbosa J, Lee K, Lee S, Lodhi B, Cho JG, Seo WK, et al. Efficient quantitative assessment of facial paralysis using iris segmentation and active contour-based key points detection with hybrid classifier. BMC Med Imaging 2016;16:23.

6) Banks CA, Jowett N, Azizzadeh B, Beurskens C, Bhama P, Borschel G, et al. Worldwide testing of the eFACE facial nerve clinician-graded scale. Plast Reconstr Surg 2017;139:491e-8e.

7) Asthana A, de la Hunty M, Dhall A, Goecke R. Facial performance transfer via deformable models and parametric correspondence. IEEE Trans Vis Comput Graph 2012;18:1511-9.

8) Kim HS, Kim SY, Kim YH, Park KS. A smartphone-based automatic diagnosis system for facial nerve palsy. Sensors (Basel) 2015;15:26756-68.

9) Mun SJ, Park KT, Kim Y, Park JH, Kim YH. Characteristics of the perception for unilateral facial nerve palsy. Eur Arch Otorhinolaryngol 2015;272:3253-9.

10) Botman JW, Jongkees LB. On transitory recovery immediately after facial nerve grafting and its cause. Acta Otolaryngol 1955;45:305-11.

11) House JW, Brackmann DE. Facial nerve grading system. Otolar-

yngol Head Neck Surg 1985;93:146-7.

12) Fattah AY, Gurusinghe AD, Gavilan J, Hadlock TA, Marcus JR, Marres H, et al. Facial nerve grading instruments: systematic review of the literature and suggestion for uniformity. Plast Reconstr Surg 2015;135:569-79.

13) O'Reilly BF, Soraghan JJ, McGrenary S, He S. Objective method of assessing and presenting the House-Brackmann and regional grades of facial palsy by production of a facogram. Otol Neurotol 2010;31: 486-91.

14) Cohen SR, Ross DA, Burstein FD, Lefaivre JF, Riski JE, Simms C. Skeletal expansion combined with soft-tissue reduction in the treatment of obstructive sleep apnea in children: physiologic results. Otolaryngol Head Neck Surg 1998;119:476-85.

15) Vrabec JT, Backous DD, Djalilian HR, Gidley PW, Leonetti JP, Marzo SJ, et al. Facial Nerve Grading System 2.0. Otolaryngol Head Neck Surg 2009;140:445-50.