Outcomes measures for patients with facial nerve injury

Garrett R. Griffin, MD[a], Jennifer C. Kim, MD[b]

From the [a]The Center for Advanced Facial Plastic Surgery, Beverly Hills, California; and
[b]Department of Otolaryngology, University of Michigan Health System, Ann Arbor, Michigan.

There are more surgical and medical options for treating patients with facial nerve injury than ever before. However, little high-quality outcomes research has been performed comparing these different interventions. Fortunately, there are a number of well-validated outcomes measures available to evaluate patients with facial nerve injury. This manuscript categorizes and summarizes these tools.
© 2012 Elsevier Inc. All rights reserved.

Facial nerve injury (FNI) can be a devastating condition with potentially severe functional, esthetic, and emotional consequences. The treatment of patients with (FNI) has always been challenging, and there are still aspects of facial nerve regeneration at the central and peripheral level that we do not fully understand. Fortunately, the past 40 years has witnessed significant advancement in the approach to the patient with FNI. For the most part, there is an accepted protocol for medical and surgical interventions, depending on the etiology and degree of injury, the amount of spontaneous facial nerve recovery, the time since injury, and the age and goals of the patient. However, within this overarching regimen, there remains significant variability in management between different physicians and centers. For example, consider a 45-year-old woman with complete facial nerve paralysis (House–Brackmann grade [HB] VI/VI) 18 months after acoustic neuroma excision and who wants to be able to generate a more symmetric smile. Two well-accepted potential interventions include temporalis tendon transfer and 2-stage free gracilis transfer. Which will achieve more dynamic movement? Which results in a better quality of life (QOL)? The ability to answer these questions depends, in part, on the existence of a valid reliable way to compare the results of these 2 interventions.

Outcome instruments are often divided into subjective and objective measures. Most would consider both provider-scored and patient self-scored evaluations to be subjec-

tive, but these 2 types of instruments produce very different information and include different biases. Thus, for this discussion, outcomes will be divided into the 3 categories of provider-scored, patient-scored, and truly objective measures. Electrophysiologic tests do provide a type of outcome measure of facial nerve function, but are discussed in a separate article. New "observer-scored" outcomes are discussed as well.

The goal of this manuscript is to compare the different outcomes measures available to assess patients with FNI. This should help practitioners to perform clinical research regarding this important population.

Provider-scored

In 1983, John House carefully reviewed the 8 most common facial nerve grading scales in use at that time[1] (Botman and Jongkees, May, Pietersen, Smith [M. Smith, personal communication, 1980], Adour and Swanson, Janssen, Yanagihara, and Stennert).[2-8] House categorized them as gross, regional, and specific scales. Gross scales sought to clump a large amount of data about the entire face into a single score. Regional scales divided the face into functional and esthetic units, typically some variation on brow, eye, and mouth regions plus or minus a midface or nasal score. In contrast, specific scales attempted to collect the maximal data. For example, the Stennert index, the only specific scale reviewed, asked a long list of yes/no questions about the face at rest and in motion, and included a separate score for

Address reprint requests and correspondence: Garrett R. Griffin, MD, The Center for Advanced Facial Plastic Surgery, Beverly Hills, CA.
    E-mail address: griffin.fpsurgery@gmail.com

secondary defects, including gustatory lacrimation ("crocodile tears") and hyperacusis. House then analyzed reliability and validity by having 15 otologists grade 12 subjects (1 normal and 11 with varying degrees of FNI) using each of the 8 scales. A complete summary is beyond the scope of this article and is excellently reviewed in another report.[9] House found that, in general, each of the scales correlated well with the others. In all but the Yanagihara scale, interobserver variability was low at both ends of the spectrum of facial function, but significantly higher in the moderate severity range. Gross scales were slightly less reliable but showed good validity and were preferred by the otologists, whereas regional scales were equally valid and more reliable. House used this information to propose a 6-interval gross facial nerve scale that was slightly altered by Brackmann to yield the House–Brackmann Grading Scale (HBGS).[10] This was adopted in 1985 by the American Academy of Otolaryngology and the American Neurotology Society as the standard for grading facial nerve function.

From that point on, the HBGS essentially replaced the other scales (except the Yanagihara scale, which is still used in Japan). The HBGS gave otologists and facial reanimative surgeons an efficient tool with which to categorize FNI, and provided a succinct language to discuss and compare these patients. The HBGS has been criticized as well. Perhaps the greatest censure has been that it lumps resting symmetry, movement, and secondary defects into a single score. Hence, a patient with HB IV/VI (moderately severe) paralysis could either have significant weakness or disfiguring contracture. These 2 hypothetical patients both have moderately severe FNI, but would look very different and would be managed very differently. A second critique of the original HBGS is that it does not allow one to accurately grade or compare patients with injury to just 1 facial nerve division or branch. For example, what is the HBGS score for a patient with excellent eye closure (HB II/VI) but severe asymmetry of the smile (HB IV/VI)?

Several modifications have been suggested in response to these criticisms. In 2003, Yen et al proposed a regional adaptation of the HBGS that separated the face into 4 regions (forehead, eye, midface, and mouth) and also included a separate synkinesis grade (none, mild, or severe).[11] They had 6 physicians score 38 patients with FNI using both the original and their modified regional HBGS. They found that the global HBGS most closely correlated with the function around the eye, and had almost no correlation to brow function. This is not surprising because House purposely weighted periocular and perioral function more than brow function in his original scale. The same group published a follow-up study in 2009 in which 14 medical student and resident physician raters scored videos of 11 FNI patients using the original global HBGS and their adapted regional version of the scale.[12] This study essentially showed worse reliability of the regional scale compared with the original global scale. Despite these findings, the Facial Nerve Disorders Committee of the American Academy of Otolaryngology proposed a House–Brackmann Facial Nerve Grading System 2.0 in 2009, which incorpo-

rated many of the same changes.[13] Facial movement was categorized into 4 regions (brow, eye, nasolabial fold, and oral) and assessed on a 5-point scale from normal to no movement. The 4 regional scores were then added together and assigned a grade from I to VI. A separate synkinesis score was also included. The HBGS 2.0 was evaluated by having 14 physicians (12 neurotologists) score videos of 21 FNI patients using both the traditional and revised HBGS 2.0. The overall interobserver reliability for the 2 scales was essentially identical. The committee concluded that the HBGS 2.0 provided additional information without decreasing reliability. Interestingly, no recommendation was made to report the regional raw data (ie, B4E2N3M3, which would correspond to brow 4/6, eye 2/6, etc), which limits the amount of new information that can be conveyed.

Two other popular physician-scored scales have been reported since the inception of the HBGS, including the Sunnybrook (or Toronto) scale and the Sydney scale, which was developed and is primarily used in Australia. The Sydney and Sunnybrook systems are relatively similar. They both include a regional assessment of movement, as well as a separate synkinesis score. Differences are that the Sydney scale deconstructs the face by facial nerve branch (including the cervical branch), whereas the Sunnybrook system grades 5 regionally focused facial movements (forehead wrinkle, gentle eye closure, open mouth smile, snarl, and lip pucker). Another difference is that the Sunnybrook system includes a score for resting symmetry, and gives a separate synkinesis score for each of the 5 assessed facial expressions. A well-designed study found significant correlation between these 2 scales (Pearson correlation coefficient ≥ 0.75) for each facial region.[14] The Sunnybrook instrument has been more extensively validated, and so will be discussed in more detail.

The authors of the Sunnybrook Facial Grading System (SFGS) sought to create a scale that would specifically address some of the perceived weaknesses of the HBGS. Most importantly, they were interested in creating a scale that was responsive to clinically significant changes in facial nerve function. By including separate component resting symmetry, voluntary movement, and synkinesis scores, as well as a weighted composite score, they hoped to create the "ideal" facial grading system. The SFGS was carefully validated from its inception. Content validity was ensured by including physicians, physiotherapists, and patients with FNI in the development of the instrument. Construct validity was assessed using the SFGS to evaluate 19 patients with FNI who had already been prospectively evaluated before and after physiotherapy with (1) a truly objective linear measurement of key facial landmarks (mentioned later in the text), (2) a detailed blinded expert visual assessment, and (3) the HBGS. The original study had demonstrated statistically significant improvements as measured by objective linear measurement and expert assessment.[15] This initial validation found that (1) the resting symmetry, voluntary movement, and synkinesis subscores varied independently, suggesting they each conveyed unique information; (2) each of the 3 subscores correlated well with the com-

posite score, meaning that the composite score weighted the subscores adequately; (3) the 3 subscores and the composite score were sensitive to clinical change, with a statistically significant improvement after therapy compared with before treatment; and (4) The HBGS did not show a statistically significant change before and after treatment. Most patients were graded III/VI both before and after a course of physiotherapy.

A follow-up study by an independent group evaluated interobserver reliability by asking 5 raters to evaluate 25 patients with FNI using the SFGS. Intraclass correlation coefficients (κ) were 0.70-0.85 (substantial agreement) for the 3 subscores and composite score.[16]

In 2006, de Ru et al[17] proposed a scale somewhat similar to the SFGS, called the Movement, Rest, Secondary Defects, and Subjective Scoring System. The subjective score is simply from 0 (no complaints) to 10 (serious complaints), and the secondary defects score included crocodile tears and hyperacusis. The scoring for this system is significantly more complicated than for the SFGS, and the added subjective score is overly simplistic and unvalidated. A PubMed search yielded no additional reports using the Movement, Rest, Secondary Defects, and Subjective Scoring System.

In summary, the HBGS and SFGS are the 2 most commonly used physician-graded facial nerve grading scales in use today. They were developed in different ways by different types of clinicians, and for different purposes. The HBGS was developed by an otologist seeking to create an informative and easy-to-use facial nerve grading system. Being a neurotologist, House was primarily interested in following facial nerve outcomes after insults occurring in the cerebellopontine angle or temporal bone (ie, the main nerve trunk). In many ways, the HBGS reflects what we know about peripheral nerve injury and regeneration. In other words, HB I-II/VI scores reflect relatively minimal nerve injury, HB III-IV/VI suggests significant injury with reasonable axonal regeneration and its attendant secondary defects, and HB V-VI/VI indicates severe injury with little to no reinnervation of the facial musculature. Succinctly, House was interested in measuring what happens to facial function after different kinds of injury to the main trunk of the facial nerve. In contrast, the SFGS was developed by physical therapists and physicians interested in accurately measuring clinical change in facial function after an intervention meant to improve that function. As such, the SFGS is a more continuous scale that collects more finite information. Both the HBGS and SFGS are valid and reliable scales that are best used in their separate intended ways.

Objective Systems

The SFGS was designed to help detect changes in facial function during and after medical and surgical interventions after FNI. One drawback is that it remains a practitioner-scored scale, and is hence subject to all the biases (particularly bias on the part of the practitioner who performed or prescribed the intervention) that afflict all subjective scales. This is one reason that there is significant interest in truly objective ways to measure facial movement. The ideal objective system would involve no human judgment (as this could be biased), be low-cost, be convenient and comfortable for the practitioner and patient, not be dependent on a motionless face, be sensitive to subtle changes, detect and measure synkinesis, and measure facial movement continuously, which would allow a calculation of both amplitude and velocity.

Noncomputerized methods

The first attempt to create an objective measure of facial movement for patients with FNI was reported by Burres[18] in 1985, called the Linear Measurement Index (LMI). He first examined 30 individuals with no facial paralysis to refine his system and tabulate the range of "normal" facial movement. Facial points that could be reliably "reidentified" (oral commissure, lateral canthus, brow at the midpupillary line) were selected, and participants were asked to make 7 standard facial expressions (soft eye closure, tight eye closure, forehead wrinkle, frown, kiss, nose wrinkle, and smile). With hand calipers, Burres measured the distance between various pairs of points (called linear measures) at rest and with facial expression, on both sides of the face. Perhaps not surprisingly, there was more variation in linear measures between 2 individuals than between the 2 sides of the face in a single person. The mean percentage of difference between measures on the 2 halves of normal individuals' faces was 5% at rest and 6% with expression, although this might be within the standard of error of this method of facial landmark determination. Burres selected the pairs of points for each expression that had the lowest ratio of variability to magnitude of movement. He then proposed an LMI, which would combine the percentage of displacement (distance between points with facial expression − distance at rest/ distance at rest) of 9 pairs of points with the 7 expressions. Burres and Fisch applied this method to patients with FNI in subsequent publications. In 1990, Croxson et al[19] used 5 observers to grade 41 FNI patients using the HBGS and the Burres–Fisch LMI, and found that in 64% of patients, the HBGS and LMI agreed completely. The LMI was more likely to grade FNI patients as having better function than as scored by the HBGS. Importantly, they found the LMI too cumbersome for routine use.

Frey et al believed that facial landmarks needed to be better classified into truly mobile and immobile points for any type of LMI to be valid. To answer this question, they used a VICON system with 4 cameras and a computer to perform 3-dimensional (3D) video analysis of normal people making various facial expressions. The VICON system (Los Angeles, CA) tracked reflective hemisphere markers that were affixed to the patient's face. They found that the "central nose point" (roughly corresponding to the rhinion) and the auricular tragi were the only points that moved <1 mm with all tested facial expressions. They then identified the pairs of facial points that best assessed movement of the

brow, eye, and nose/mouth complex (please refer to primary article for full details). This system worked but was time-consuming and expensive. Hence, they developed Frey's "faciometer," a digital caliper that could be used to measure these facial distances in the clinic. Although this did save some time and expense, a single patient still required 20 minutes for data collection. Other drawbacks of the faciometer were that it could only evaluate 1 facial expression at a time, and could not assess synkinesis or velocity of movement. Still, Frey's analysis of facial landmarks is an important contribution to the field of facial movement analysis.

In 2008, Manktelow et al reported a way to use 2 plastic rulers in the clinic to measure outcomes of facial reanimation of the perioral region.[20] They validated the ruler system using 2 raters to evaluate the resting mouth position of 21 FNI patients, as well as the dynamic movement in 10 normal individuals. Intra- and interrater reliability (by intraclass correlation coefficients) was ≥0.88 (excellent) for both the static and dynamic measures. The dynamic measurements in the normal patients were further analyzed for accuracy by comparing the ruler measurements with those determined by a computerized system (discussed further later in the text). The average difference between the ruler and computer measurements was 1.7 mm (~10% of the actual commissure excursion in healthy individuals), which was considered adequately low.

Computerized methods

Most modern methods of objective facial analysis use a computer in 1 of 2 ways. In the first approach, the landmark facial points are identified with some type of bead or mark. A video is made of the patient performing facial expressions, which are then analyzed on a computer to measure movement of the selected points. In a second technique, no markers are placed on the face. The patient is videotaped making facial expressions, and a different type of computer program is used to identify small regions of the face that have different color or reflectance (essentially, pixels that have changed) to quantify movement.

Marker-based systems

Johnson et al in 1994 were the first to report computerized analysis of facial points as a way to measure facial movement.[21] Small adhesive dots were applied to the face at key points, and a piece of a metric ruler was attached to the nasal tip such that it did not obscure any of the marks. The patients were then photographed making 5 expressions: maximal brow lift, tight eye closure, maximal smile, maximal frown, and maximal whistle/pucker. The photographs were projected onto a grid in such a way that 1 cm on the ruler in the projected image equaled 1 cm on the grid. Absolute movement in the vertical (y) and horizontal (x) dimensions of the dots with the 5 expressions was then calculated for 7 normal subjects and 3 patients with FNI. The typical magnitude along the primary plane of movement (x or y) was 0.4-0.9 cm. There was significant be-

tween-subject variation for the magnitude of movement. The method was decently reliable, with an intrasubject between-measure standard deviation of 0.07 cm, or approximately 10% the magnitude of movement. One advantage of this technique is that it can measure synkinesis by assessing movement of points distant from the facial expression (ie, oral commissure movement during tight eye closure). One drawback is that this method measured maximal movement only; hence, velocity of movement is not readily attainable.

Isono et al[21] reported a similar technique using 24 facial markers; 4 points were midline, with 10 lateralized points on each side of the face. They videotaped 44 normal subjects and 12 acute FNI patients with Bell palsy or Ramsay–Hunt syndrome during eye closure. The video was fed into a computer, and movement of the 10 lateralized points on each side of the face was measured, although the authors do not explain how movement in the images was translated into an accurate metric measurement (mm). The ratio of total movement of the 10 points on the paretic side to the normal side was used as an outcome, and was shown to gradually increase as patients with acute FNI regained facial movement. The authors are to be commended for assessing a fairly large group of normal subjects to establish a "normal" range of symmetry (≥85%). Weaknesses of their report are that there is no explanation of how they selected the 24 dots, or what anatomic points were used to ensure reliable placement between measures; indeed, several points were placed on the lateral cheek, where there are no fixed anatomic reference points. The authors did not measure velocity but discuss tracking points at 10 separate frames throughout the motion, which would enable an approximation of velocity. They also report no reliability assessment.

Tomat and Manktelow[22] modified the marker-based technique in 2005 to assess dynamic smile reanimation. A water-soluble pencil was used to mark 8 points around the mouth at the red lip/philtral junction, at the oral commissures, at the midline of the lower-lip vermilion border, and at the bilateral upper- and lower-lip midlip points, which are halfway between the midline and commissure. The central nose point between the tip-defining points was used as a reference to control for facial motion, and a metric ruler was taped to the chin to establish a scale. Twenty patients with FNI were evaluated by 2 raters, on 2 occasions, 2 weeks apart. At each of the 2 time points, frontal, left, and right three-quarter views were videotaped at rest and with maximal smile. The purpose of the three-quarter video was to view the mouth in the plane of motion of the smile. The authors show that the camera need only be within 15° of this plane to achieve ≤3% variation in measurement. The video was fed into a computer, and the rest and maximal smile frames were isolated and overlapped at 50% saturation, using the central nose point to align the images. This overlapped image was then uploaded into Adobe Photoshop (Adobe, San Jose, CA), and movement of the dots was assessed using the ruler on the chin to allow measurement in millimeters. Importantly, the angle of movement of the points was assessed as well. Interrater reliability for distance and angle measurements was excellent at >0.98. The intra-

rater reliability was >0.95 for both distance and angle. The accuracy of the system was assessed by using it to quantify distances and angles that had already been measured using a microcaliper; the system differed from the caliper measurement by 4% for distance and 2° for angle, on average. This system has excellent reliability and could be used for other facial expressions. It is time-consuming (20 minutes per patient) but reasonable for research purposes. It does not measure velocity. Synkinesis was not assessed but could be if more facial markers are applied and analyzed.

Kang et al[23] used an almost identical technique that same year, but used a full-profile view instead of a three-quarter view. This allowed them to calculate movement in the x, y, and z (anterior–posterior) planes, thus enabling transformation to a true 3D vector of movement with distance and angle components. These authors used the ocular light reflex to correct for any head movement on frontal view, and used the auricle to overlap profile views. The authors used this system to assess the movement of 8 perioral points in 50 healthy Korean adults, thus providing a range of normal values for distance and angle in this population. The commissure was the perioral point that moved the most: an average of 13.7 mm (range, 6-22 mm) at an average angle of 48.9° (range, 30°-62°) to horizontal. The authors did not break this range of movements down by gender, but this underscores the variability in facial movement between normal individuals.

There have been some studies of the differences obtained from a 2-dimensional (2D) versus 3D measurement of facial movement. Gross et al[24] in 1996 used a computerized marker-based system to analyze movement of various facial markers in 2 dimensions using a frontal camera compared with the 3D measurement achieved using a frontal and 2 lateral cameras, which enabled inclusion of anterior–posterior movement. They examined movement of commissure, Cupid's bow, alar rim, supraorbital, infraorbital, and chin points during smile, cheek puff, eye closure, grimace, and lip purse. The 2D measurement was less than the 3D measurement for all points. The magnitude of this difference was proportional to the amount of movement. As an example, the 3D measurement was 2.9 mm greater than the 2D one for oral commissure excursion with smile.

In his 2002 Triologic thesis, Linstrom[9] used the Peak Motus (VICON, Los Angeles, CA) to simultaneously track reflective markers placed on the rhinion, nasion, upper eyelids, and oral commissures. Thirty-four normal subjects and 26 FNI patients were analyzed. Participants rested their chin on a cushion and were videotaped during gentle eye closure and closed-mouth smile. Three trials of each expression were completed. The video data were uploaded into the Peak system, and once a particular point was selected, its movement was tracked throughout all subsequent frames. This system continually tracks a point through time, allowing velocity calculations. Primary outcomes were the percentage of asymmetry between the 2 sides of the face, and movement in centimeters. Synkinesis was defined as movement of a point distant from the site of interest greater than the 95th percentile in normal subjects. The system had 95%

sensitivity in detecting movement abnormalities, and noted abnormalities in 40% of FNI patients graded as HB I/VI. Excessively large (>90th percentile of normal) movement occurred in approximately 30% of FNI patients on the normal side during both eye closure and smile; this reflects that FNI patients can develop hyperkinesis on their intact side in an attempt to recruit all nearby motor units. Synkinetic eye movements were found in >50% of the FNI patients, and 40% of patients had synkinetic smiles. Test–retest reliability of the percentage asymmetry measure was >0.7 for normal and paretic patients, and was actually higher in the FNI patients. The only drawback of the system is that at the time of publication in 2002, the analysis took 25-45 minutes for each patient. The Peak Motus system is still for sale (http://www.vicon.com).

Automated Facial Analysis (AFA) is a method originally designed to recognize emotional expression, which uses several "computer vision" approaches to quantify facial motion. In this system, facial landmarks (ie, the oral commissure) can be selected with a mouse and will be continuously monitored from that point forward; thus, no marks need to be placed on the patient's face. AFA was adapted by Wachtman et al[25] in 2001 to objectively quantify facial movement. To prove validity, dots were placed on 9 FNI patients' face as described by Johnson, and the patients were videotaped performing brow raise, eye closure, and smile. Facial movement was then also analyzed using AFA and compared with the results of movement analysis using Johnson's maximal static response assay (discussed later in the text). Pearson correlation coefficient between the 2 measurement systems was >0.95 for all measures; on average, the 2 systems differed by just 0.2-0.3 mm. The AFA system has 3 advantages over many of the marker-based systems. Firstly, the head can move within 5° of rotation without compromising the results. Secondly, no markers need to be applied to the patient's face, which saves time and patient comfort. Finally, the system tracks the selected points continuously allowing velocity measurements.

Figure 1 demonstrates the most common marker locations used in this type of objective facial movement analysis.

Pixel-/luminance-based systems

The second broad category of facial movement analysis uses computers to count pixels or changes in light reflectance in various regions of the face. Neely et al[26] were the first to use this technique. They took black-and-white video of 5 normal subjects and 17 patients with facial paralysis during 5 voluntary maneuvers: rest, forehead wrinkle, eye closure, nose wrinkling, and mouth smiling. The video was digitized onto a computer, and images during movement were "subtracted" from the image at rest. Pixels that did not change their brightness would cancel becoming dark, whereas facial regions that moved would have different brightness between the 2 images and would be light. A particular "threshold" was then selected to turn the light and dark gray pixels in the combined image into black (no or

Figure 1   The most commonly used facial landmarks for objective facial movement analysis: (1) superior brow at midpupillary line, (2) upper lid lash-line at midpupillary line, (3) lower lid lash-line at midpupillary line, (4) rhinion (relatively immobile), (5) "central nose point" between tip-defining points (relatively immobile), (6) oral commissure, (7) mid–upper-lip points, (8) midline upper and lower lips, and (9) mid–lower-lip points. (Color version of figure is available online.)

little movement) and white (a lot of movement) pixels that could be counted. This was done for images 100 ms apart through the facial movements, allowing the creation of "strength–duration" curves showing movement over time during an expression, a corollary of velocity. This technique also allowed the assessment of synkinesis, which would be seen as changes of luminosity in areas of the face distant from the desired movement. This method correlated highly with the HBGS, suggesting validity, but no reliability testing was performed. The major drawback of this technique is that the head must be completely still during the videotaping for the subtraction method to be accurate. In a later publication, the group assessed the validity and reliability of their system for assessing brow elevation, eye closure, and smile.[27] They chose a different outcome than pixel number, instead settling on the ratio of the left- to right-side pixel counts in normal individuals, and ratio of the paretic- to normal-side pixels in FNI patients. In 38 normal subjects, the mean and median values were very close to 1.0, or perfect symmetry, for all 3 expressions. Brow elevation revealed the most asymmetry, with the 10%-90%-isle range going from 0.85 to 1.10. There were outliers in brow elevation with ratios of 0.75 and 1.45, showing that even among normal patients, there can be significant facial asymmetry. Test–retest reliability for the system was assessed using 30 FNI patients with varying degrees of facial function. The intraclass correlation coefficients for all 3 expres-

sions were >0.95 (excellent). Pixel ratios exceeded 1.0 for brow elevation, eye closure, and smile, reflecting hyperkinesis and spasm in some of these FNI patients. Synkinesis was not specifically addressed in this study, but is measurable with this technique.

Meier-Gallati and Fisch[28] developed a very similar technique called OSCAR (Objective SCaling of facial nerve function based on ARea analysis). They used the term luminance to describe their outcome measure of facial analysis, but their overall approach is very similar to Neely's. They evaluated 14 normal subjects and 20 patients with FNI using this technique. Reliability was not assessed. A confusing attempt was made to validate the method by comparing scores with the HBGS. Only rest and maximal movement images were analyzed, making velocity of movement impossible to assess. Approximately 10% of the measurements had to be discarded because patient head movement made the analysis unreliable. No mention of synkinesis was made.

He et al[29] proposed a different computational approach to achieve the same ends. This system uses artificial neural networks to determine pixel changes on the 2 halves of the face, and computes a ratio of these changes on the paretic side compared with the intact side. Thus, the "outcome" is a value typically between 0 and 1, although synkinesis and spasm can raise this value above 1. This system, later termed the Glasgow Facial Palsy Scale, was validated by comparing it with the HBGS, Yanagihara scale, and SFGS.[30] Forty FNI patients were videotaped making 5 standardized expressions: raise eyebrow, close eyes gently, close eyes tightly, wrinkle nose, and full smile. The videos were uploaded to a laptop with the Glasgow Facial Palsy Scale software and analyzed. For validation purposes, 3 otolaryngologists independently scored the videos using the 3 facial grading scales. Correlation coefficients between the computerized Glasgow scale and the 3 subjective scales were between 0.64 (HBGS) and 0.72 (Yanagihara scale), which is moderately strong. The strengths of this system are that it is automated and objective. The software could be shared between centers, requiring just a video camera and laptop for setup. Regional data, and hence synkinesis, assessment is possible, but was not performed in the validation study. This system is designed primarily to assess maximal movement, although it could theoretically be modified to collect data throughout a facial expression, yielding velocity data. The primary weakness of the system is that no discrete measurement of facial movement, in terms of distance or angle, is achieved.

Several groups interested in dynamic smile reanimation have recently reported outcomes using simple post hoc analysis of clinic photographs at rest and with smile. Two different techniques have been used to establish a metric scale in photographs. Some centers have used the intercanthal distance to establish a scale in the photomanagement package MIRROR (Canfield, Fairfield, NJ).[31,32] Another group has used the relatively constant size of the adult human iris (11.8 mm, standard deviation = 0.42 mm) to set a scale. The Massachussetts Eye and Ear Infirmary used the
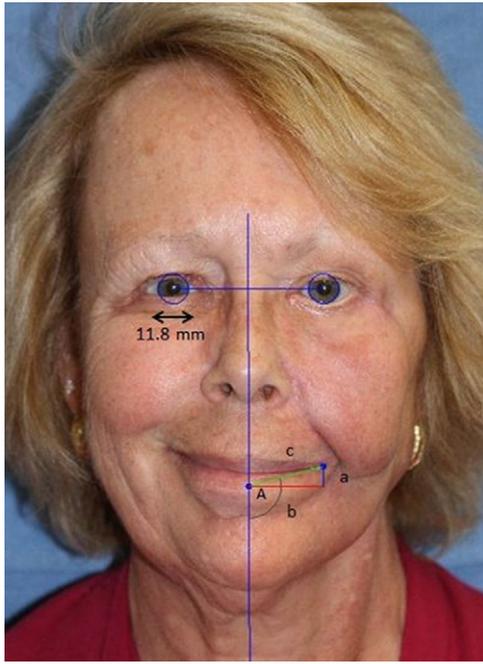
Figure 2 The Scaled Measurement of Improvement in Lip Excursion software program. The user manually identifies the cornea by choosing 4 points on the iris margin. The program automatically connects the irises with a horizontal line, and drops a perfectly perpendicular vertical midline. The user then chooses facial landmarks to analyze. For example, for oral commissure movement, the midline lower lip and commissure points are selected. The software will automatically calculate the distances a, b, and c as well as the angle A. By comparing these values at rest and with smile, excursion can be calculated. Movement of the brow, midlip points, and philtrum can be assessed using this system as well. (Color version of figure is available online.)

iris technique to create the Scaled Measurement of Improvement in Lip Excursion system.[33] This requires one to define the size of the iris in photographs, after which the program automatically calibrates a distance scale, draws a line between the irises, and drops a vertical midline. The system can then calculate distances on the face. By comparing these distances in pictures at rest and with smile, excursion can be determined. Our group recently used this software to assess resting symmetry and excursion after temporalis tendon transfer, and found it to be very user friendly.[34] One drawback of this system is that it measures excursion in 2 dimensions only, losing the anterior–posterior component of commissure excursion. Also, velocity calculations are not possible. Figure 2 demonstrates the basic technique.

Table 1 compares the various objective facial movement analysis systems.

Considerations

There is currently no standard method of obtaining objective facial movement data. Many authors have envisioned a "facegram" resembling an audiogram that could compare facial movement on the intact and paretic sides with a "normal standard." One obstacle for this approach is that there is significant variation in facial movement between normal individuals. Any normal standard would need to be based on age, gender, and race to be of any value. Many studies discussed earlier in the text used the intact side of the face as standard for comparison. This is logical, but studies by Linstrom[9] and others have shown that the intact side of patients with FNI can become significantly hyperkinetic. Thus, the "normal" side becomes abnormal itself and is a less reliable standard of comparison. One can sometimes demonstrate this by covering the paretic side in photographs of FNI patients; when viewed in isolation, it becomes clear that the unaffected side often has also become "deformed" owing to hyperkinesis (Figure 3).

Patient-scored subjective

In recent years, there has been an increased focus on health-related QOL (HRQL) data. This is partially in response to rising health care costs and the need to understand the value that various interventions provide to patients and society. Both generic and disease-specific QOL measures are available for the FNI population.

General Scales

Several studies have used well-validated nonspecific QOL measures to assess FNI. The Short Form Health Survey (SF-36) is a general questionnaire developed for use in health economics and in the calculation of quality-adjusted life years.[37] Its 36 items are segregated into 8 domains, many of which are relevant to patients with FNI: vitality, physical functioning, bodily pain, general health perceptions, physical role functioning, emotional role functioning, social role functioning, and mental health. The format of answer choices varies throughout the questionnaire. Numerous studies have shown the validity and reliability to be good or excellent.

The Glasgow Benefit Inventory is an 18-item questionnaire designed specifically to assess health changes after otolaryngologic procedures (ie, surgery).[38] All questions require the patient to compare their postintervention with their preintervention state. Answers are scored on a 5-point Likert scale from 1 (much worse) to 5 (much better). Validity was originally demonstrated by showing significant changes in groups of patients after tonsillectomy, middle ear surgery, and rhinoplasty. To our knowledge, reliability has never been assessed.

The Derriford Appearance Scale was developed to assess psychological distress and dysfunction resulting from esthetic problems of appearance.[39] It has good validity and reliability and is highly sensitive to changes resulting from operative intervention. It comes in a 59-question standard format with 5 subscales (general self-consciousness of appearance, social self-consciousness of appearance, negative self-concept, sexual and bodily self-consciousness of appearance, and facial self-consciousness of appearance), as well as a 24-item short form.

Table 1   Comparison of objective facial analysis systems

Author	Type (marker-based, etc.)	Name of technique	What is measured?	Format of results (output)	Head motionless?	Measures synkinesis?	Measure velocity?
Burres[18]	Handheld caliper	Linear Measurement Index	Distance between facial points	Sum of movement over 7 expressions	No	No	No
Frey[35]	Marker-based computerized	VICON	Distance between facial points	Percentage of maximal distance between points	No	Yes	Possible, not done
Frey[35]	Digital caliper	Faciometer	Distance between facial points	Actual change in distance between points (in mm)	No	No	No
Manktelow et al[20]	Rulers	Handheld ruler	Distance between facial points	Actual change in distance between points (in mm)	No	No	No
Johnson et al[36]	Marker-based computerized	Maximal Static Response Assay	Actual movement of facial points	Actual movement of facial points along x- and y-axis (in mm)	No	Yes	No
Isono et al[21]	Marker-based computerized	None	Actual movement of facial points	Movement of facial points along x- and y-axis (in mm) Comparison of movement between sides of the face	No	Yes	Possible, not done
Tomat and Manktelow[22]	Marker-based computerized	Facial Reanimation Measurement System	Actual movement of facial points	Actual linear movement of points (in mm)	No	Possible, not done	No
Kang et al[23]	Marker-based computerized	None	Actual movement of facial points	Actual linear movement of points (in mm)	No	Possible, not done	No
Linstrom[9]	Marker-based computerized	Peak Motus	Actual movement of facial points	Actual linear movement of points Percentage of asymmetry between sides of the face	Yes	Yes	Yes
Wachtman et al[25]	Landmark-based computerized	Automated Facial Analysis	Movement of facial points over time	Pixels of movement (can be converted to distance in mm) Visual path of movement over time	No, 5° of movement tolerated	Yes	Yes
Neely et al[26,27]	Pixel-based computerized	Facial Analysis Computerized Evaluation	Changes in facial luminance	Number of pixels that change (area or mm²) Ratio of changes between sides of the face	Yes	Yes	Yes
Meier-Gallati et al[28]	Luminance-based computerized	OSCAR	Changes in facial luminance	Percentage (area or mm²) of a region of the face that changes	Yes	Yes	Possible, not done
Kecskes et al[30]	Pixel-based computerized	Glasgow Facial Palsy Scale	Changes in pixel characteristics	Ratio of changes between sides of the face	No	Possible, not done	No
Bray and Hadlock et al[33]	Landmark-based	Scaled Measurement of Improvement in Lip Excursion	Distance between facial points	Actual change in distance between points (in mm)	No	Yes	No

Specific scales

The Facial Disability Index (FDI) was reported by the University of Pittsburgh group in 1996 as a disease-specific HRQL instrument designed to evaluate FNI patients.[40] Its 10 items are divided into physical function and social/well-being function subscales, with 5 questions each. The authors have attempted to establish validity and reliability, although this could have been done in a more convincing manner.

The Facial Clinimetric Evaluation (FaCE) Scale is a 15-item questionnaire with a global score and 6 subscales for facial movement, facial comfort, oral function, eye comfort, lacrimal control, and social function. The authors asked 86 FNI patients in 2 separate geographic areas of the United States to complete the FaCE Scale, the FDI, and the SF-36.[41] Two weeks later, these patients completed the FaCE Scale again to allow test–retest reliability determination. Test–retest reliability overall and for the subscales was high (Spearman $r > 0.8$, $P < 0.01$). Scores on the initial FaCE, FDI, and SF-36 for the 86 subjects were compared with those of 23 normal individuals. There was a significant difference in overall FaCE and FaCE subscale scores, as well as both FDI domains, between the FNI and normal groups, suggesting construct validity. There was no difference in SF-36 domain scores between the 2 groups, suggesting that the SF-36 is not adequately sensitive to detect important differences in HRQL related to facial nerve function. Forty-one of the FNI patients returned to the clinic and were scored using the HBGS and SFGS. Interestingly, the physician-scored synkinesis component of the SFGS did not correlate well with any of the FaCE subscales. This likely occurred because the FaCE scale includes items about facial tightness, discomfort, and fatigue, but there are no specific questions about synkinesis.

In response to this deficiency, Mehta et al[42] developed and validated the Synkinesis Assessment Questionnaire. It consists of 9 items answered on a 5-point Likert scale, with higher scores representing worse synkinesis. Test–retest reliability was high (Spearman $r = 0.88$, $P < 0.0001$). The difference in scores between FNI patients and normal control subjects was statistically significant, as was the difference in scores for FNI patients before and after chemodenervation, demonstrating construct validity.
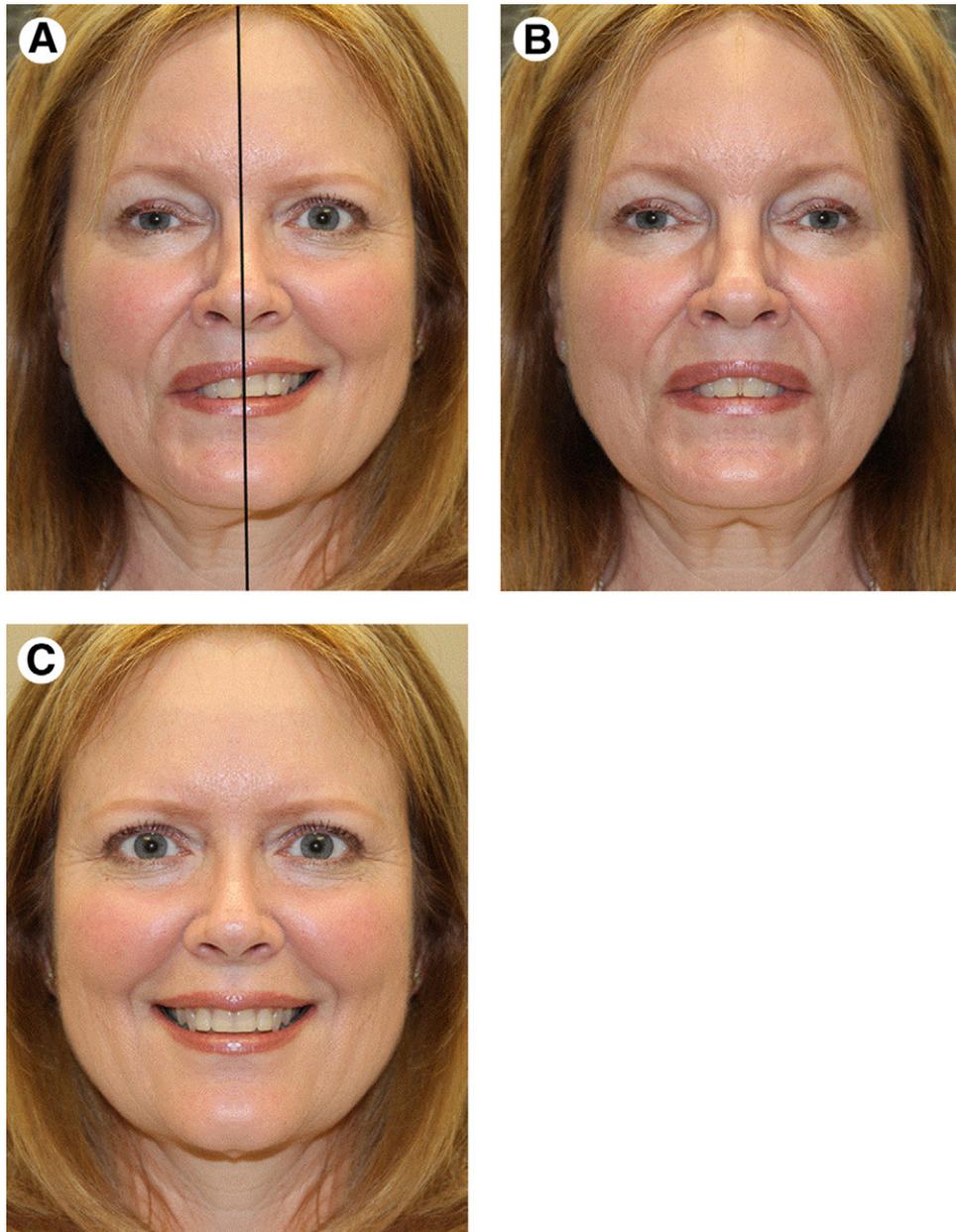
Figure 3 Demonstration of synkinesis and hyperkinesis. (A) 52 year-old woman who suffered right-sided Bell palsy during pregnancy at age 29. She has developed significant synkinesis on the right side and some compensatory hyperkinesis on the left side. (B) The affected right side has been mirrored showing that she clearly cannot communicate a smile on that side. (C) The intact left side has been mirrored showing clear communication of a smile, but in an extreme hyperkinetic manner. (Color version of figure is available online.)

Deviation of the nasal base, with resultant nasal obstruction, is an underappreciated result of FNI. The Nasal Obstruction Symptom Evaluation scale is a 5-item questionnaire using a 5-point Likert scale.[43] It is has good reliability and validity, and has been used in several studies to assess patients with a variety of rhinologic complaints.[44]

de Almeida et al[45] recently developed the Lip Reanimation Outcomes Questionnaire. It is a 15-item instrument with 7 questions asking about the severity of preanimation lip/mouth function and 8 questions asking the patient to compare their postoperative with their preoperative state. Answers are scored on a 7-category Likert scale, except for item 15, which uses a visual analog scale. The authors also included a 4-item provider-scored section. Validation was poorly performed in a small population of patients. Hence, other outcomes instruments should be considered until more thorough validation has occurred.

Future directions

Provider-scored, patient-scored, and objective outcomes essentially tell us how caregivers think the patient is doing, how the patient thinks the patient is doing, and how the computer thinks the patient is doing, respectively. However, what about the patient's husband, boss, or gardener?

The Johns Hopkins group has recently published 2 fascinating studies using observers to evaluate facial function.

The first study sought to identify how much asymmetry was tolerable in the brow and oral commissure regions before it became recognizable. Computer-manipulated photographs with progressively increasing amounts of asymmetry were shown to blinded observers. The authors found that at least 3 mm of asymmetry in both regions was required before it was noticeable. When observers spent more time looking at the photographs, progressively less asymmetry was required for detection. When both the brow and oral commissure were altered, asymmetry around the eye was identified more quickly than around the mouth; however, with longer viewing times, both were noticed.[46]

In a related experiment, Ishii et al examined whether patients with FNI were able to clearly express their emotions. Blinded observers viewed photographs of normal individuals and FNI patients and completed a questionnaire asking them to identify the affect being displayed (happy, disgust, anger, sadness, and fear), as well as the observer's own subjective judgment about the person (trustworthy, friendly, neutral, hostile, energetic, tired). Photographs of FNI patients smiling were evaluated as negative 73% of the time compared with just 2% for the control subjects (Figure 3). Overall, patients with facial paralysis were seen as displaying a negative affect the majority of the time.[47]

Discussion

There are many ways to measure facial function in patients with FNI. The array of outcomes instruments is interesting because both otologists and reanimative surgeons have developed methods tailored to their specific needs. There is no single perfect outcome instrument, even among subjective or objective techniques alone. In most cases, the ability to collect more precise data requires more time, complexity, and money.

Providers without the luxury of significant ancillary support should still be able to collect good outcomes data if they are organized. At our center, we typically use the original HBGS to score patients at their initial visit for documentation purposes. We do not currently use the SFGS, but it is an excellent provider-scored instrument. All FNI patients complete the FaCE and Nasal Obstruction Symptom Evaluation questionnaires in the waiting room before being seen. Quickly reviewing these scales at the beginning of the visit allows the physician to hone in on anything that is bothering the patient. Currently, there is no agreed-on standard for truly objective outcomes instruments, but most modern techniques are based on analysis of photographs or video of standard facial expressions. Photographic stills are easily pulled from video using readily available software like iMovie (Apple, Cupertino, CA) or Windows Live Movie Maker (Microsoft, Redmond, WA), so separate photos are not absolutely necessary. Being diligent about obtaining video and/or photographs at each visit ensures that at least some objective facial movement data can be obtained at a later date.

Recent advancements in computers, cameras, and facial recognition software should allow continued progress toward an affordable, quick, and automated objective assessment of facial motion. It is not hard to imagine a future in which FNI patients obtain a "facegram" from an audiologist or nurse before seeing their physician. Recent studies quantifying how much facial asymmetry is noticeable, and whether emotions can be clearly conveyed, are truly fascinating. Eliminating distracting features and enabling patients to accurately convey their intended emotions may become part of the standard of success in treating patients with FNI.

References

1. House JW: Facial nerve grading systems. Laryngoscope 93:1056-1069, 1983
2. Botman JW, Jongkees LB: The results of intratemporal treatment of facial palsy. Pract Otorhinolaryngol 17:80-100, 1956
3. May M: Facial paralysis, peripheral type: A proposed method of reporting. (Emphasis on diagnosis and prognosis, as well as electrical and chorda tympani nerve testing) Laryngoscope 80:331-390, 1970
4. Pietersen E: Spontaneous course of Bell's palsy, in Fisch U (ed): Facial Nerve Surgery, Proceedings: Third International Symposium on Facial Nerve Surgery, Zurich, 1976. Amstelveen, The Netherlands, Kugler Medical Publications, 1977, pp 337-343
5. Adour KK, Swanson PJ Jr: Facial paralysis in 403 consecutive patients: emphasis on treatment response in patients with Bell's palsy. Trans Am Acad Opthlamol Otolaryngol 75:1284-1301, 1971
6. Janssen FP: Over de postoperatieve facialis-verlamming [Thesis]. Amsterdam, the Netherlands, University of Amsterdam, 1963
7. Yanagihara N: Grading of facial palsy, in Fisch U (ed): Facial Nerve Surgery, Proceedings: Third International, Symposium on Facial Nerve Surgery, Zurich, 1976. Amstelveen, The Netherlands, Kugler Medical Publications, 1977, pp 533-535
8. Stennert E: Facial nerve paralysis scoring system, in Fisch U (ed): Facial Nerve Surgery, Proceedings: Third International Symposium on Facial Nerve Surgery, Zurich, 1976. Amstelveen, The Netherlands, Kugler Medical Publications, 1977, pp 543-547
9. Linstrom CJ: Objective facial motion analysis in patients with facial nerve dysfunction. Laryngoscope 112:1129-1147, 2002
10. House JW, Brackmann DE: Facial nerve grading system. Otolaryngol Head Neck Surg 93:146-147, 1985
11. Yen TL, Driscoll CL, Lalwani AK: Significance of house-Brackmann facial nerve grading global score in the setting of differential facial nerve function. Otol Neurotol 24:118-122, 2003
12. Reitzen SD, Babb JS, Lalwani AK: Significance and reliability of the house-Brackmann grading system for regional facial nerve function. Otolaryngol Head Neck Surg 140:154-158, 2009
13. Facial Nerve Disorders Committee, Vrabec JT, Backous DD, et al: Facial nerve grading system 2.0. Otolaryngol Head Neck Surg 140:445-450, 2009
14. Coulson SE, Croxson GR, Adams RD, et al: Reliability of the "Sydney," "Sunnybrook," and "House-Brackmann" facial grading systems to assess voluntary movement and synkinesis after facial nerve paralysis. Otolaryngol Head Neck Surg 132:543-549, 2005
15. Ross BG, Fradet G, Nedzelski JM: Development of a sensitive clinical facial grading system. Otolaryngol Head Neck Surg 114:380-386, 1996
16. Kayhan FT, Zurakowski D, Rauch SD: Toronto facial grading system: interobserver reliability. Otolaryngol Head Neck Surg 122:212-215, 2000
17. de Ru JA, Braunius WW, van Benthem PP, et al: Grading facial nerve function: why a new grading system, the MoReSS, should be proposed. Otol Neurotol 27:1030-1036, 2006
18. Burres SA: Facial biomechanics: the standards of normal. Laryngoscope 95:708-714, 1985

19. Croxson G, May M, Mester SJ: Grading facial nerve function: house-Brackmann versus Burres-Fisch methods. Am J Otol 11:240-246, 1990

20. Manktelow RT, Zuker RM, Tomat LR: Facial paralysis measurement with a handheld ruler. Plast Reconstr Surg 121:435-443, 2008

21. Isono M, Murata K, Tanaka H, et al: An objective evaluation method for facial mimic motion. Otolaryngol Head Neck Surg 114:27-31, 1996

22. Tomat LR, Manktelow RT: Evaluation of a new measurement tool for facial paralysis reconstruction. Plast Reconstr Surg 115:696-704, 2005

23. Kang YS, Bae YC, Hwang SM, et al: A simple and quantitative method for three-dimensional measurement of normal smiles. Ann Plast Surg 54:379-383, 2005

24. Gross MM, Trotman CA, Moffatt KS: A comparison of three-dimensional and two-dimensional analyses of facial motion. Angle Orthod 66:189-194, 1996

25. Wachtman GS, Cohn JF, VanSwearingen JM, et al: Automated tracking of facial features in patients with facial neuromuscular dysfunction. Plast Reconstr Surg 107:1124-1133, 2001

26. Neely JG, Cheung JY, Wood M, et al: Computerized quantitative dynamic analysis of facial motion in the paralyzed and synkinetic face. Am J Otol 13:97-107, 1992

27. Neely JG, Wang KX, Shapland CA, et al: Computerized objective measurement of facial motion: normal variation and test-retest reliability. Otol Neurotol 31:1488-1492, 2010

28. Meier-Gallati V, Scriba H, Fisch U: Objective scaling of facial nerve function based on area analysis (Oscar). Otolaryngol Head Neck Surg 118:545-550, 1998

29. He S, Soraghan JJ, O'Reilly BF, et al: Quantitative analysis of facial paralysis using local binary patterns in biomedical videos. IEEE Trans Biomed Eng 56:1864-1870, 2009

30. Kecskés G, Jóri J, O'Reilly BF, et al: Clinical assessment of a new computerised objective method of measuring facial palsy. Clin Otolaryngol 36:313-319, 2011

31. Byrne PJ, Kim M, Boahene K, et al: Temporalis tendon transfer as part of a comprehensive approach to facial reanimation. Arch Facial Plast Surg 9:234-241, 2007

32. Sidle DM, Fishman AJ: Modification of the orthodromic temporalis tendon transfer technique for reanimation of the paralyzed face. Otolaryngol Head Neck Surg. 145:18-23, 2011

33. Bray D, Henstrom DK, Cheney ML, et al: Assessing outcomes in facial reanimation: evaluation and validation of the SMILE system for measuring lip excursion during smiling. Arch Facial Plast Surg 12:352-354, 2010

34. Griffin GR, Abuzeid W, Vainshtein J, et al: Outcomes following temporalis tendon transfer in irradiated patients. Arch Facial Plastic Surg (in press)

35. Frey M, Jenny A, Giovanoli P, et al: Development of a new documentation system for facial movements as a basis for the international registry for neuromuscular reconstruction in the face. Plast Reconstr Surg 93:1334-1349, 1994

36. Johnson PC, Brown H, Kuzon WM Jr, et al: Simultaneous quantitation of facial movements: the maximal static response assay of facial nerve function. Ann Plast Surg 32:171-179, 1994

37. Hays RD, Sherbourne CD, Mazel RM: The rand 36-item health survey 1.0. Health Econ 2:217-227, 1993

38. Robinson K, Gatehouse S, Browning GG: Measuring patient benefit from otorhinolaryngological surgery and therapy. Ann Otol Rhinol Laryngol 105:415-422, 1996

39. Harris DL, Carr AT: The Derriford appearance scale (DAS59): a new psychometric scale for the evaluation of patients with disfigurements and aesthetic problems of appearance. Br J Plast Surg 54:216-222, 2001

40. VanSwearingen JM, Brach JS: The facial disability index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. Phys Ther 76:1288-1300, 1996

41. Kahn JB, Gliklich RE, Boyev KP, et al: Validation of a patient-graded instrument for facial nerve paralysis: the FaCE scale. Laryngoscope 111:387-398, 2001

42. Mehta RP, WernickRobinson M, Hadlock TA: Validation of the synkinesis assessment questionnaire. Laryngoscope 117:923-926, 2007

43. Stewart MG, Witsell DL, Smith TL, et al: Development and validation of the nasal obstruction symptom evaluation (NOSE) scale. Otolaryngol Head Neck Surg 130:157-163, 2004

44. Most SP: Analysis of outcomes after functional rhinoplasty using a disease-specific quality-of-life instrument. Arch Facial Plast Surg 8:306-309, 2006

45. de Almeida JR, Alexander AJ, Shrime MG, et al: Development and preliminary validation of the lip reanimation outcomes questionnaire. Otolaryngol Head Neck Surg 143:361-366, 2010

46. Chu EA, Farrag TY, Ishii LE, et al: Threshold of visual perception of facial asymmetry in a facial paralysis model. Arch Facial Plast Surg 13:14-19, 2011

47. Ishii LE, Godoy A, Encarnacion CO, et al: What faces reveal: impaired affect display in facial paralysis. Laryngoscope 121:1138-1143, 2011