Clinician-Graded Electronic Facial Paralysis Assessment: The eFACE

Caroline A. Banks, M.D.
Prabhat K. Bhama, M.D., M.P.H.
Jong Park, B.S.
Charles R. Hadlock, Ph.D.
Tessa A. Hadlock, M.D.
Boston and Waltham, Mass.; and
Anchorage, Alaska

Background: The subjective nature of facial aesthetics and the difficulties associated with quantifying facial function have made outcomes analysis in facial paralysis challenging. Clinicians rely on photographs, subjective descriptions, and scales, limiting assessment, communication among providers, and communication between providers and patients. The authors describe the development and validation of a comprehensive, electronic, clinician-graded facial function scale (eFACE), which generates an overall disfigurement score and offers simple graphic output for clinician communication, assessment of various interventions, and patient understanding. The eFACE application may be used in a variety of electronic devices, including smartphones, tablets, and computers.

Methods: An instrument consisting of 16 items in a visual analogue scale format was developed to assess facial function and symmetry (the eFACE). Video recordings of subjects performing facial expressions were viewed, and the eFACE instrument was applied, along with an overall facial disfigurement score. A multiple regression analysis was performed to determine the best linear relationship between overall expert-determined disfigurement and the eFACE items. The resulting equation was tested by three independent facial nerve clinicians, using an additional series of patients, to determine both interrater and intrarater reliability of the instrument.

Results: Multiple regression analysis produced good fit of eFACE parameters to overall expert-rated global facial disfigurement when dynamic parameters were weighted twice as heavily as static and synkinesis parameters. eFACE scores demonstrated very high interrater and intrarater reliability.

Conclusion: The eFACE is a reliable, reproducible, and straightforward digital clinical measure with which to assess facial function and disfigurement in patients with facial paralysis. (Plast. Reconstr. Surg. 136: 223e, 2015.)

Comprehensive, objective assessment of facial harmony and function in patients suffering from facial movement disorders has eluded clinicians for decades. In the modern era of evidence-based practice, quantitative, rigorously validated outcomes measures are essential to gauge severity of disease, to monitor progression over time, and to determine the effectiveness of different interventions. To date, physicians who treat these conditions often rely on subjective descriptions and crude scales of facial anatomical features and function, which precludes efficient communication with other physicians and severely hampers the advancement of treatment techniques by preventing statistical analysis of outcomes. Even among facial reanimation experts, parameters

Disclosure: None of the authors has a financial interest in any of the products or devices mentioned in this article.

Supplemental digital content is available for this article. Direct URL citations appear in the text; simply type the URL address into any Web browser to access this content. Clickable links to the material are provided in the HTML text of this article on the Journal’s Web site (www.PRSJournal.com).
and outcomes are often recorded using nonvalidated and/or subjective descriptions in the medical record, and supplemented with nonstandard photographs. Although photographs and videos provide an excellent record of patient progress, they do not provide an efficient means by which outcomes can be objectively compared over time or between individuals, and remain inadequate for direct patient communication.

Our objective in this study was to leverage our experience in examining all different levels and states of abnormal position and movement in patients suffering from facial movement disorders, to create an instrument for assessment of facial function and symmetry in patients with facial paralysis, and to mathematically correlate the score to overall disfigurement. Our goal was to design a scale that is easy to use, uses smartphone or desktop/laptop technology, produces a visually interpretable output, and provides data relevant to all practitioners involved in facial paralysis management. A universal, visually meaningful scale of facial function, permitting all practitioners involved to clearly understand the progression or resolution of disease, is crucial for prognosis, referral, treatment planning, and assessment of long-term outcomes. Although clinician-graded scales for patients with facial paralysis exist, they are limited in their ability to comprehensively analyze static, dynamic, and synkinetic facial features independently; are not universally adopted; and do not generally provide a graphic or pictorial means by which the patient can effectively understand the assessment. Equally importantly, no existing scale has been mathematically correlated to overall disfigurement; this project was designed to address this deficiency by first establishing an equation relating disfigurement to the clinician-graded scale, and then testing the intrarater and interrater consistency of scores generated by applying the equation to the scale scores generated by a group of facial nerve practitioners.

**METHODS**

**Item and Scale Generation**

An observer-graded visual analogue scale was designed and tested in this study. The scale (termed the eFACE) is a 16-item instrument relevant to patients with facial nerve disorders. The items on the scale were chosen based on both in-depth knowledge and experience with existing clinician-rated scales and our institution’s deep experience in treating patients with facial paralysis. Similar to several other scales in use, the items were designated into one of three domains: static (five items), dynamic (seven items), or synkinesis (four items), based on the constructs that they addressed. Static items included resting brow height, resting palpebral fissure width, nasolabial fold depth at rest, nasolabial fold orientation at rest, and oral commissure position at rest. Dynamic items included brow excursion with attempts at elevation, palpebral fissure narrowing with attempts at gentle eye closure, palpebral fissure narrowing with attempts at full eye closure, oral commissure movement with smile, nasolabial fold depth with smile, nasolabial fold orientation with smile, and lower lip movement. Synkinesis items included ocular synkinesis, midfacial synkinesis, mentalis dimpling, and platysmal synkinesis. [See Figure, Supplemental Digital Content 1, which shows the eFACE, a 16-item, digital instrument that assesses three parameters: static (five items), dynamic (seven items), and synkinesis (four items), http://links.lww.com/PRS/B353.]

For each parameter, a value of 0 indicated the most extreme malposition (static items), total absence of movement (dynamic items), or the most severe synkinesis (synkinesis items). A value of 100 indicated balanced position (static items), normal movement (dynamic items), or absence of synkinesis (synkinesis items). For parameters in which the pathologic state can be manifest in two directions, the scale was expanded from 0 to 200, and the absolute value of the distance from 100 was used for score computation. Examples of such parameters include resting brow position, which may be ptotic (score <100), balanced (score of 100) or paradoxically hyperelevated (score >100); and nasolabial fold depth and orientation, which can be effaced and/or vertically oriented, or can be hyperprominent and/or horizontally oriented.

Scores were then calculated, by simple averaging, to produce a subscore for each domain of facial function (static, dynamic, and synkinesis); subscores ranged from 1 to 100, with 100 representing normal facial symmetry and/or function. The software automatically generated a graphic display of each item in the format of a bar chart (Fig. 1). The software also has the ability to generate a graphic display of serial facial nerve examinations, thus documenting progression of the facial movement disorder over time (Fig. 2). From our data repository, a series of patient photographs representing the full range of dysfunction in each facial zone was used to develop a template for reference and teaching purposes (see Figure, Supplemental Digital Content 2, which shows an instructional template...
of patient photographs representing the full range of dysfunction in each facial zone, http://links.lww.com/PRS/B354, and to develop a list of rules and exceptions (Table 1), with the idea that clinicians may refer to the template or a brief instructional video as a primer before using the scale. (See Video, Supplemental Digital Content 3, which shows the eFACE tutorial video demonstrating use of the eFACE grading scale along with examples of facial paralysis, ranging from mild to severe, available in the “Related Videos” section of the full-text article on PRSJournal.com or, for Ovid users, available at http://links.lww.com/PRS/B355.)

Disfigurement Rating

A simple continuous disfigurement scale was used, similar to previously published disfigurement ratings,26–28 in which a value of 100 indicated the normal state, and a value of 1 indicated extreme disfigurement.

Study Population and Grading Procedures

Seventy-four subjects with varying degrees of facial paralysis were selected randomly from our practice. A single surgeon with extensive experience in examining the paralyzed face (at least 5000 in-person examinations of patients representing the entire spectrum of flaccid through hypertonic states) viewed brief video recordings of each subject performing a series of standardized facial expressions: repose, eyebrow raise, gentle eye closure, full eye closure, small smile, full smile, pursing the lips, depressing the lower lip, counting from 1 to 3, and saying “happy birthday.” A digital 16-item eFACE assessment and an overall facial disfigurement score were assigned to each patient.

For eFACE scoring, all static assessments were made from a paused still frame of the video clip, after which the video was run, and paused as needed to enter the scoring digitally on the sliding scales. For synkinesis assessments, the entire 30-second clip was rerun, with particular attention paid to the four zones affected by synkinesis. Synkinesis severity was always judged by the movement that produced the most severe synkinesis. For example, if there was mild involuntary narrowing of the palpebral fissure width during smiling but severe palpebral fissure width narrowing during puckering, the latter score would be recorded. For disfigurement scoring, the entire video was observed, after which a disfigurement rating was assigned.

Model Development

A multiple regression analysis was performed for the overall disfigurement rating against the composite eFACE scores for the three categories (static, dynamic, and synkinetic), using
Fig. 2. Serial facial nerve examinations in a single individual, who traverses from the flaccid, through improved, to a synkinetic state. (Above) Serial bar graphs demonstrate the progression of facial dysfunction. (Below, left) Flaccid facial paralysis after acute insult to the facial nerve (blue bars). (Below, center) Nearly balanced facial nerve examination with ongoing recovery (black bars). (Below, right) Severe synkinesis and hypertonicity following spontaneous recovery (green bars). NLF, nasolabial fold.
the 74-subject sample. Once the best linear relationship between expert-determined disfigurement and eFACE scores was established, both intrarater reliability and interrater reliability were tested in a new series of 25 patients. Three independent experienced facial nerve clinicians each applied the eFACE scale two times for each subject, exactly as described above, with a 2-week interval between the independent trials.

**RESULTS**

**Assessments**

Expert-rated assessments of facial function using the eFACE scale took on average 90 seconds per patient, after an initial learning curve in which assessments took up to 4 minutes. Disfigurement scores took approximately 60 seconds each. The eFACE scores did not differ significantly when data were entered during direct patient contact, or when recorded from video clips after the patient interaction (data not shown).

**Demographics of the Test Population**

An independent group of 25 subjects were chosen for assessments of interrater and intrarater reliability. A majority (60 percent) of the subjects were women. The mean age was 55 years, with a range from 27 to 84 years (Table 2). The most common cause in the cohort was Bell palsy, followed by vestibular schwannoma, and others (Table 3).

**Multivariable Linear Regression**

Multivariable linear regression of the clinician-graded disfigurement severity on the three subscale scores yielded coefficients of −0.012, −0.029, and −0.011 for the static, dynamic, and synkinesis subscales, respectively, thus corresponding to the model equation: severity = 6.73−0.012*static−0.029*dynamic−0.011*synkinesis. This equation was determined to show excellent agreement with the clinician’s estimate, yielding a coefficient of determination (\(R^2\)) value of 0.811, with the analysis of variance statistic and all individual coefficients being highly significant (\(p < 0.01\)). Multicollinearity was insignificant.

**Intrarater and Interrater Reliability**

We tested intrarater reliability (repeatability) for each of the three raters separately by applying a paired \(t\) test to their repeated assessments of the 25 subjects. In all three cases, there was no significant
difference between the set of ratings, even at a con-
servative 10 percent level of significance. For interra-
ter reliability, we used the first of the two assessments
by each of the raters because this would best repre-
sent the situation of a typical rater, who would only
be conducting a single assessment. Figure 3 shows
the overall severity values calculated by our model
from the 16 individual values assigned by each rater
for each of the subjects. This agreement is seen
to be very strong over the entire range of severity
levels. The calculated intraclass correlation
coefficient for these values is 0.97, based on a two-
way random effects model, often referred to as
intraclass correlation coefficient.29,30 Here I would
represent perfect conformity, and thus this is
confirmation of excellent reliability.

DISCUSSION

Facial paralysis is a notoriously difficult prob-
lem to describe, rate, measure, and follow lon-
gitudinally. In current practice, the emerging
trend is to perform comprehensive assessments31
that include clinician-graded scales, patient-
reported outcomes and quality-of-life instru-
ments, objective or quantitative measures of
function, and even layperson assessments of dis-
figurement.27 Unfortunately, consensus among
facial nerve practitioners regarding appropriate
scales for clinician-graded and patient-reported
scales is lacking.1 Even though the American
Academy of Otolaryngology–Head and Neck
Surgery adopted the House-Brackmann scale
as a standard facial nerve rating scale, the scale
was intended solely to describe recovery follow-
ing the extirpation of vestibular schwannomas,
and thus is often erroneously applied to a wide
range of other pathologic conditions that result
in facial paralysis. In addition, the interrater
and intrarater reliability is only fair,32–34 and
a second-generation House-Brackmann scale
(House-Brackmann II)5,34 designed to address

Table 2. Demographics and Patient Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>15 (60)</td>
</tr>
<tr>
<td>Male</td>
<td>10 (40)</td>
</tr>
<tr>
<td>Age, yr</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>55.04 ± 14.56</td>
</tr>
<tr>
<td>Range</td>
<td>27–84</td>
</tr>
<tr>
<td>Time, mo*</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>47 ± 124.8</td>
</tr>
<tr>
<td>Range</td>
<td>0.03–588</td>
</tr>
</tbody>
</table>

*Approximate length of time from onset of symptoms to presenta-
tion at the Facial Nerve Center.

Table 3. Cause of Facial Paralysis

<table>
<thead>
<tr>
<th>Pathologic Condition</th>
<th>No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bell palsy</td>
<td>11 (44)</td>
</tr>
<tr>
<td>Vestibular schwannoma</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Geniculate ganglion hemangioma</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Iatrogenic</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Trauma</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Facial nerve neuroma</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Ramsey Hunt syndrome</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Lyme</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Parotid gland malignancy</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Idiopathic</td>
<td>1 (4)</td>
</tr>
</tbody>
</table>

Fig. 3. Overall severity values calculated from the 16 individual values assigned by each rater for
each of the subjects. Note that there is excellent agreement over the entire range of severity levels.
the shortcomings of the original scale has not been widely adopted. A recent review of all clinician-rated scales reveals that, to date, there still exists no ideal method of rapid, effective assessment to permit communication among providers, and between providers and patients.

The effectiveness of a graph or picture in communicating both among providers and in the context of provider-patient discussions cannot be overstressed. Their utility has been elegantly addressed in otology by the clinical audiogram, which represents an excellent example of how the development of an objective, visual assessment of a construct can lead to universal understanding of disease severity among clinicians. The audiogram has served as a foundation for hearing loss treatments, hearing outcomes research, and as a means by which the clinician can visually communicate a complex physiologic construct to the patient and other providers, enhancing both patient understanding and facilitating multidisciplinary care. In this article, we attempted to provide a similarly useful tool for facial paralysis rating.

In the present work, we developed and tested a 16-item visual analogue scale for facial assessment, which provides numeric and graphic outputs for each item, and computes static, dynamic, and synkinesis subscores. In addition, by establishing the relationship between the eFACE subscores and overall expert-determined degree of disfigurement, we have produced a weighted output reflective of overall facial function and appearance. We fully acknowledge that disfigurement is a subjective and elusive construct to measure, and the data in this article use the disfigurement scores of a single facial nerve clinician. However, there does exist some advantage of using the perceptions of clinicians with daily, extensive exposure to the entire spectrum of facial movement disorders for the development of disfigurement scales, because they are in general more familiar with the breadth of the spectrum of disfigurement caused by facial paralysis.

We have demonstrated high interrater and intrarater correlations, exceeding those of the current House-Brackmann standard, and have designed the graphic output to be clear, comparative, and easy to use in direct patient care. A visual analogue scale was used as opposed to a simple descriptive ordinal scale, to optimize sensitivity for detecting differences. The instrument is efficient, and can be administered by means of smartphone, tablet, laptop, desktop, or on paper, and transferred later to an electronic system for subscore and total score computations and graphic outputs. Any equipment with touch-screen technology allows extremely rapid assessment, although using a mouse for input did not result in significant slowdown in data entry.

The eFACE scale described in this article provides important static, dynamic, and synkinesis information, and a global facial function score. The subscores in these three areas may represent a powerful tool for assessing the effects of specific interventions. For example, the effect of a surgical intervention designed only to improve static position of facial landmarks may not be detectable using existing scales, which lack such sensitivity. Thus, following brow ptosis correction or the placement of an eyelid weight, the eFACE possesses the potential to register an improvement in the static subscore, whereas more global or less specific clinician-rated scales would overlook the benefit achieved. Likewise, as patients emerge from flaccid facial paralysis and experience the onset of synkinesis, the eFACE might demonstrate stable overall disfigurement scores, but will accurately reflect the significant changes in static, dynamic, and synkinesis subscores, to guide changing therapy.

Electronic zonally based clinician-rated facial function scales add an elegant new tool to overall assessment of facial function in patients suffering from facial movement disorders. Shortcomings of the tool involve its lack of utility in patients with bilateral facial paralysis, and the fact that continuous scales, in exchange for higher sensitivity to small changes, require more upfront “learning” to establish definitions of different levels of function. Moreover, variability and bias of clinician perceptions represents an important weakness of this and other similar clinician-rated scales, and the multivariable linear regression coefficients were derived from single-surgeon ratings of disfigurement.

Despite the fact that the eFACE is easy to use, its use would represent a paradigm shift in the documentation of facial function toward an electronic record and, without the backing of the facial nerve community, is likely to suffer major obstacles to widespread adoption. In addition, although the current report presents development of the instrument, critical follow-on work is necessary to provide concrete examples of the utility of the instrument in actual clinical scenarios. To this end, our ongoing work involves applying eFACE assessment to prerecovery and postrecovery states, and to preintervention and postintervention states, to demonstrate its applicability and determine its sensitivity to changing clinical pictures. Furthermore, a large cohort of patients must be examined prospectively, by multiple facial nerve clinicians from multiple institutions and training backgrounds, to
further refine the weighting coefficients for each of the 16 items, and ultimately to achieve an optimal fit to clinician assessments of disfigurement, using a continuous disfigurement scale.

Despite the fact that a great deal of study is required to define the precise utility of the eFACE instrument in clinical practice, its ease of utility and its digital and database-amenable format provide an attractive platform for ongoing studies.

Caroline A. Banks, M.D.
Division of Facial Plastic and Reconstructive Surgery
Department of Otolaryngology/Head and Neck Surgery
Harvard Medical School/Massachusetts Eye and Ear Infirmary
243 Charles Street
Boston, Mass. 02114
caroline_banks@meei.harvard.edu

PATIENT CONSENT
Patients provided written consent for the use of their images.

REFERENCES