Sunnybrook Facial Grading System: Intra-rater and Inter-rater Variabilities
Camille Cabrol, Léa Elarouti, Anne-Laure Montava, Sylvie Jarze, Julien Mancini, Jean-Pierre Lavieille, Pauline Barry, Marion Montava

To cite this version:

HAL Id: inserm-03650974
https://www.hal.inserm.fr/inserm-03650974
Submitted on 13 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
TITLE: SUNNYBROOK FACIAL GRADING SYSTEM: INTRA-RATER AND INTER-RATER VARIABILITIES

Short Running Head: Variabilities of Sunnybrook facial grading system

Authors: CABROL Camille 1*, camcabrol@gmail.com, Speech therapist
ELAROUTI Léa* 1*, lea.elarouti@outlook.fr, Speech therapist
MONTAVA Anne-Laure 1, anne-laure.montava@wanadoo.fr, Speech therapist
JARZE Sylvie 1, sylvie.jarze@icloud.com, Speech therapist
MANCINI Julien 2,3, julien.mancini@ap-hm.fr, MD PhD
LAVIEILLE Jean-Pierre 1,4, jean-pierre.lavieille@ap-hm.fr, MD PhD
BARRY Pauline 1, pauline.barry@ap-hm.fr, MD
MONTAVA Marion 1,5, marion.montava@ap-hm.fr, MD PhD

Affiliations:
1: Department of Otorhinolaryngology-Head and Neck Surgery, Conception Hospital, Aix-Marseille University, 147 Boulevard Baille, F-13005, Marseille, France
2: Department of Biostatistics, Timone Hospital, Aix-Marseille University, 264 Rue Saint Pierre, F-13005, Marseille, France
3: Aix Marseille Univ, INSERM, IRD, APHM, UMR1252, SESSTIM, Department of Public Health (BIOSTIC), Hôpital de la Timone, Marseille, F-13005, France
4: Aix Marseille University, IFSSTAR, LBA, UMR-T 24, F-13344, Marseille, France
5: Aix Marseille University, CNRS, INP, UMR 7051, Marseille, France

* C.C. and E.L. contributed equally.

The authors do not have any conflict of interest or financial disclosure to declare.

Corresponding Author:
Marion MONTAVA, Department of Otorhinolaryngology-Head and Neck Surgery, Conception Hospital, Aix-Marseille University, 147 Boulevard Baille, 13005 Marseille, France. Telephone: +33 491 435 520; Fax: +33 491 435 419; e-mail address: marion.montava@ap-hm.fr
Sunnybrook Facial Grading System: Intra-rater and inter-rater variabilities

ABSTRACT:

Objectives: Evaluate intra-rater and inter-rater variabilities of the Sunnybrook Facial Grading System (SFGS) and identify potential factors of variability.

Study design: Prospective test of hypothesis.

Setting: University tertiary referral centre.

Participants/Methods: Facial video recordings of 20 patients with variable degrees of peripheral facial palsy (PFP) were anonymized then randomly presented to 31 independents raters in two trials. The raters were senior and junior professionals involved in the management of PFP: ENT specialists, physiotherapists and speech therapists. The SFGS was used for grading paralyses.

Main outcome measure: Intra-rater and inter-rater variabilities were estimated by intraclass correlation coefficient (ICC [95% confidence interval]) for the composite score and the three subscores of the SFGS. Factors of variability studied were: rater professions and rater experience (senior vs junior).

Results: For the total population, the intra-rater ICC was 0.915 [0.900-0.929] for the composite score considered to represent almost perfect repeatability. Repeatability was important for symmetry at rest (0.694 [0.646-0.737]), almost perfect for voluntary movements (0.903 [0.886-0.918]) and synkinesis (0.810 [0.778-0.838]). The inter-rater ICC for the composite score was 0.847 [0.755-0.923] indicating almost perfect agreement between all raters. Agreement between raters was almost perfect for voluntary movements (0.839 [0.746-0.919]), but moderate for symmetry at rest (0.553 [0.408-0.730]) and synkinesis (0.476 [0.333-0.666]). Some differences were found between raters groups; however, repeatability and agreement were good for all raters.
Conclusions: The SFGS is a reproducible scale. It can be used with good reproducibility by both novices and experts, and by all professionals involved in the management of PFP.

KEYWORDS:
Facial palsy,
Facial grading scales,
Sunnybrook Facial Grading System,
Inter-rater variability,
Intra-rater variability.
INTRODUCTION

Peripheral facial palsy (PFP) is a pathology caused by a facial nerve injury which affects the motricity of the hemiface on the side of the lesion. Clinicians need an objective, reliable and reproducible clinical tool to accurately describe motor facial function, assess the severity of paralysis, its evolution over time and the effects of treatments.

To this day, a multitude of different facial nerve grading instruments were developed which shows that none of them is perfect and how difficult it is to assess PFP. In 2015, a systematic review on the facial function grading instruments found that the Sunnybrook Facial Grading System (SFGS) was the scale which best accomplishes the goals of assessment and recommended its widespread adoption as the current standard in reporting outcomes of facial nerve disorders (1). Indeed, the House and Brackmann Grading system is the best known and most widely used system for its general ease of use. However, it is not giving specific details about facial function.

Introduced by Ross in 1996 (2), SFGS assesses facial resting symmetry compared to normal side, symmetry of voluntary movement and potential synkinesis associated with specified voluntary movement. Three subscores are obtained giving a composite score from 0 for complete PFP to 100 for normal facial function. However, the SFGS remains a subjective scale which makes it subject to limitation because of varying degrees of variability. From 2000 to 2010, some studies on SFGS’s variability concluded that SFGS was generally reliable (3–7). Nevertheless, they found contradictory results of variability for individual scores.

The main objective of our study was to evaluate intra-rater and inter-rater variabilities of the SFGS. Secondly, we tried to identify potential factors of variability as the profession or the level of experience in the management of patients with PFP.
MATERIALS AND METHODS

Ethical considerations

The study protocol was assessed according to guidelines of the national committee on research involving human subjects and was approved by Institutional Review Board (Ethics committee of Aix Marseille University 2019-17-10-001).

Study protocol

Subject assembly was possible through access to the facial nerve centre perpetual database comported video archive of subjects with PFP. This resource is under the control of the last author and protected behind a two-lock secure system.

After statistical consultation and review of literature, it was arbitrarily determined that 20 subjects would be sufficient to test the variability of the scale. The inclusion criteria were: good-quality video recording, unilateral PFP, all five facial voluntary expressions necessary for SFGS scoring present, and a wide range and gradated distribution of facial movement among the 20 subjects’ video images. Subjects with facial trauma were excluded. Diagnosis, time course in recovery, age, gender, or race were not considered criteria for selection, and the videotapes were anonymized. The selection of videotapes and the randomised order of viewing were determined by a committee of three experts.

A group of 31 independent raters was constituted by senior and junior professionals involved in the management of patients with PFP: ENT specialists, physiotherapists and speech therapists. The raters analysed the videotapes and graded the facial function of each subject using the SFGS and the HBGS in 2 independent and timed rounds with a 2-months interval. Neither documentation nor training session were provided. During session, the raters could pause or go back at any time if needed to have a better analysis as long as the chronometer was
not stopped. The time of evaluation was reported. Moreover, information about their experience in the management of patients with PFP and a signed consent were reported.

Statistical analysis

To estimate intra-rater and inter-rater variabilities, the intraclass correlation coefficient (ICC [95% confidence interval]) was estimated (8). ICC was estimated for the composite score, the three subscores and each movement for the total population and for each group of raters.

According to Landis Koch (Biometrics, 1977), we considered the agreement as weak if rated within 0-0.40, moderate within 0.41-0.60, important within 0.61-0.80 and almost perfect within 0.81-0.99. Analysis were carried out with SPSS 20.0 for Windows.

Inter-rater variability was studied in the first session in order to avoid a learning effect even if the 2-months interval should be sufficient for this.

As ICC can reflect the degree of agreement but also relate to the number and the variability among the sampled observations, we did not compare statistically different measurements or raters. However, if the confidence intervals of the ICC were different without overlap between two groups, the reliability was considered to be different, better or worse.

RESULTS

In the first session, 31 raters participated: 6 ENT specialists (4 seniors and 2 juniors), 4 physiotherapists (3 seniors and 1 junior) and 21 speech therapists (9 seniors and 12 juniors). In the second session, 25 of them (80.6%) took part: 5 ENT specialists (3 seniors and 2 juniors), 3 physiotherapists (2 seniors and 1 junior) and 17 speech therapists (6 seniors and 11 juniors).
The average time of evaluation was 64.9 minutes (min) for both sessions with a standard deviation of 22.8. It was 70.9 ± 23.4 min for session 1, and 57.5 ± 19.8 min for session 2.

The information gathering highlighted the following points: all juniors had had a former PFP teaching, however, only 3/15 (20%) of them had already used the SFGS; and 12/16 (75%) of senior had at least one training course on PFP but only 9 (56.3%) used SFGS in their clinical practice.

Intra-rater variability (repeatability)

Results of intra-rater variability for the composite score and the three subscores are reported in Table 1. The composite score ICC for the total population was 0.915 [0.900-0.929] considered to represent almost perfect repeatability. Repeatability was important for symmetry at rest (0.694 [0.646-0.737]), almost perfect for voluntary movement (0.903 [0.886-0.918]) and synkinesis (0.810 [0.778-0.838]).

Results of intra-rater variability for the five standard expressions in voluntary movement and synkinesis are reported Table 2. For voluntary movement, agreement was almost perfect for the five movements performed. For synkinesis, repeatability was moderate for the gentle eye closure (0.600 [0.541-0.653]) while for the four other movements it was important.

No difference of repeatability was found between our raters groups of seniors and juniors, except for resting symmetry with better repeatability for seniors (0.791 [0.736-0.836]) than for juniors (0.620 [0.543-0.687]) and synkinesis with again better repeatability for seniors (0.880 [0.847-0.907]) than for juniors (0.726 [0.665-0.777]).

No difference was found between our raters groups of professions, except for the composite score and voluntary movement with better repeatability for ENT specialists and speech therapists than for physiotherapists.
For extreme HB grades (II and VI), repeatability was almost perfect for the composite score (0.962 [0.947-0.972]) and the voluntary movement (0.961 [0.946-0.972]), important for resting symmetry (0.723 [0.636-0.792]) and moderate for synkinesis (0.576 [0.458-0.674]). For medium HB grades (III, IV, V), repeatability was almost perfect for the composite score (0.836 [0.802-0.865]) and the voluntary movement (0.828 [0.792-0.859]) and important for resting symmetry (0.677 [0.616-0.730]) and for synkinesis (0.786 [0.742-0.823]). Repeatability was better for the composite score and voluntary movement of extreme HB grades than medium HB grades. It was worse for synkinesis and there was no difference for resting symmetry.

Inter-rater variability (agreement between raters)

Results of inter-rater variability for the composite score and the three subscores are reported in Table 3. The ICC for the composite score was 0.847 [0.755-0.923] indicating almost perfect agreement between all raters. Agreement was almost perfect for voluntary movement (0.839 [0.746-0.919]), but moderate for symmetry at rest (0.553 [0.408-0.730]) and synkinesis (0.476 [0.333-0.666]).

Results of inter-rater variability for the five standard expressions in voluntary movement and synkinesis are reported Table 2. For voluntary movement, agreement was almost perfect for forehead wrinkle (0.805 [0.701-0.899]) and open mouth smile (0.816 [0.714-0.905]). It was important for gentle eye closure (0.724 [0.588-0.852]), snarl (0.754 [0.629-0.870]) and lip pucker (0.611 [0.467-0.774]). For synkinesis, agreement was weak for forehead wrinkle (0.387 [0.258-0.582]), gentle eye closure (0.300 [0.188-0.488]), open mouth smile (0.373 [0.246-0.568]) and snarl (0.340 [0.219-0.532]). It was moderated for lip pucker (0.439 [0.302-0.632]). No differences between seniors and juniors raters groups were reported. No difference was found between our raters groups of professions, except for voluntary movement where physiotherapists had worse agreement than ENT specialists and speech therapists.
For extreme HB grades (II and VI), agreement was almost perfect for the composite score (0.948 [0.873-0.991]) and the voluntary movement (0.954 [0.887-0.992]), important for resting symmetry (0.738 [0.511-0.945]) and weak for synkinesis (0.018 [-0.004-0.169]). For medium HB grades (III, IV, V), agreement was important for the composite score (0.691 [0.529-0.855]) and the voluntary movements (0.704 [0.546-0.863]), moderate for resting symmetry (0.461 [0.299-0.696]) and for synkinesis (0.427 [0.269-0.666]). Repeatability of extreme HB grades was better than medium HB grades for the composite score and voluntary movement, and worse for synkinesis. There was no difference for resting symmetry.

**DISCUSSION**

SFGS assesses resting symmetry compared to normal side, symmetry of voluntary movement and potential synkinesis associated with specified voluntary movement. Resting symmetry is assessed by a comparison to the normal side of the palpebral fissure (normal, narrow, wide), the naso-labial fold (normal, absent, less or more pronounced), and the corner of the mouth (normal, drooped or pulled up/out). The rating is done through a points-giving system (0, 1, 2). Then, different regions of the face are examined separately, with five standard expressions used to assess the symmetry of voluntary movement and the degree of synkinesis associated with movement. The five standard expressions reflect the motor function of the five peripheral branches of the facial nerve: forehead wrinkle (frontalis), gentle eye closure (orbicularis oculi), open mouth smile (zygomaticus and risorius), snarl (levator labii superioris alaeque nasi and levator labii superioris) and lip pucker (orbicularis oris superior and inferior). The symmetry of voluntary movement for each standard expression is graded on a five-points scale from 1 (no movement) to 5 (movement complete), depending on the degree of muscle excursion compared to normal side. The degree of synkinesis associated with each standard expression is rated on a
four-points scale from 0 (no synkinesis) to 3 (severe synkinesis). Three scores are obtained and weighted as follows: the resting symmetry score is multiplied by five, and the voluntary movement score is multiplied by four. Then, a composite score is calculated by subtracting the resting symmetry score, and the synkinesis score from the voluntary movement score. A composite score of 100 corresponds to normal facial function and a composite score of 0 to complete PFP.

As it is a scale administered by rater with their own clinical experience, the SFGS remains a subjective tool subject to limitation. Previous studies on SFGS’s variability had some similar and contradictory outcomes (3–7). Our purpose was to evaluate intra- and inter-rater variabilities of the SFGS in a new study with more patients and more raters. To go further, we decided to include raters from different professions and experiences. The statistical method chosen to evaluate intra-rater and inter-rater variabilities can have a marked effect on study outcome. The choice of ICC was coherent with those previous studies and then allowed a comparison of outcomes.

**Intra-rater variability (repeatability)**

In our study with 31 raters, repeatability evaluated by ICC for the composite score was 0.900 to 0.929. ICC varied from 0.838 to 0.929 with eight novice assessors in Hu et al. (4), 0.864 to 0.995 with 26 doctors in Kanerva et al. (6), 0.948 to 0.970 for 2 naïve raters in Neely et al. (7). These results are considered to represent almost perfect repeatability.

Although the repeatability score for resting symmetry was considered as important (0.694 [0.646-0.737]) in the current study, it was almost perfect for Kanerva et al. (0.841 [0.500-0.976]). However, Kanerva et al. specified that results with coefficient of repeatability (CR) varied for resting symmetry: repeatability was only moderate or fair.
We found an almost perfect repeatability for the voluntary movement score (0.903 [0.886-0.918]) and for each movement performed. This result is consistent with the findings of Kanerva et al. (0.918 [0.799-0.988]).

In our study, repeatability for the synkinesis was almost perfect (0.810 [0.778-0.838]). This result approaches the good findings of Kanerva et al. (0.979 [0.931-0.998]).

No difference of repeatability was found between our raters groups of seniors and juniors for the composite score and voluntary movement. This result is congruent with Hu et al. and Kanerva et al. SFGS can be used as reliably by both experts and novice users. However, contrary to Hu et al. and Kanerva et al., resting symmetry and synkinesis shown better repeatability for seniors than for juniors which suggests that resting symmetry and synkinesis evaluation requires experience.

No difference was found between our raters groups of professions, except for composite score and voluntary movement with better reproducibility for ENT specialists and speech therapists than for physiotherapists. It suggests that SFGS repeatability is the same for all professionals involved in the management of PFP, except for physiotherapists. This outcome has to be treated cautiously owing to the heterogeneous sizes of the raters groups included in our study.

Thus, the results of the intra-rater variability enable to conclude that the SFGS is a reproducible scale from one time to another.

**Inter-rater variability (agreement between raters)**

In our study with 31 raters, agreement evaluated by ICC for the composite score was 0.847 [0.755-0.923]. It was 0.885 [0.76-0.92] in Kayhan et al. (3) with 5 ENT specialist, 0.892 in Hu et al. with 2 novice users, 0.997 [0.992-1.000] in Kanerva et al. with 26 doctors, 0.890 [0.784-0.946] in Neely et al. with 2 naive raters. These results represent an almost perfect agreement between raters.
We found a moderate agreement for symmetry at rest (0.563 [0.408-0.730]). However, Hu et al. found an almost perfect agreement (0.950) as Kanerva et al. (0.983 [0.960-0.996]). Kayhan et al. found an important agreement (0.72 [0.58-0.84]). In the current study, agreement between raters was almost perfect for voluntary movement (0.839 [0.746-0.919]). This outcome is congruent with Kayhan et al. (0.83 [0.73-0.90]), Hu et al. (0.976) and Kanerva et al. (0.997 [0.992-0.999]). Coulson et al. (5), with 6 ENT specialists, found important agreement (0.63).

Agreement was almost found perfect for forehead wrinkle and open mouth smile, and important for gentle eye closure, snarl and lip pucker. Neely et al. specified that forehead wrinkle and lip pucker were most variable and significantly different from the other movements.

For synkinesis we found moderated agreement (0.476 [0.333-0.666]). It was almost perfect in Hu et al. (0.913) and Kanerva et al. (0.987 [0.969-0.997]), but they didn't find as good results using CR to assess agreement: it was, only moderate or fair. Agreement was important for Kayhan et al. (0.70 [0.55-0.93]). Coulson et al. found a weak agreement (0.23).

In our study, agreement was weak for forehead wrinkle, gentle eye closure, open mouth smile and snarl, and moderate for lip pucker. These results are opposite to Kayhan et al. who found an important agreement for each movement except frontal synkinesis which showed a weak agreement (0.38 [0.21-0.59]). Thus, agreement between raters for synkinesis was highly variable. Those discrepancies can be due to the fact that the rater has to observe the whole face, whereas for voluntary movement the rater’s attention is directed to a specific region of the face.

It should be noted that in some previous studies (3,4,6,7), raters had preliminary training before the first session which may have helped to reach higher results of agreement with SFGS than ours.

Differences between seniors and juniors raters groups were not significant. This result is consistent with Hu et al. and Kanerva et al. and suggest that the SFGS can be used with little prior knowledge of the scale and then as reliably by both experts and novice users.
No difference was found between our raters groups of professions, except for voluntary movement with better repeatability for ENT specialists and speech therapists than for physiotherapists. It suggests that agreement on SFGS is the same for all professionals involved in the management of PFP, except for physiotherapists. Again, this outcome has to be treated cautiously owing to the heterogeneous sizes of the raters groups.

Thus, the results of the inter-rater variability enable to conclude that the SFGS is a reproducible scale from one rater to another. To reduce some ambiguities due to subjective assessment, Neely and al suggested criteria for completion of the SFGS.

Time of scoring

We reported the unequal times of scoring between session 1 (70.9 min) and session 2 (57.5 min). Knowing that a few raters use the SFGS in their practice (20% of juniors, 56.3% of seniors), we could suggest a more efficient using of the scale during session 2. It wouldn’t be likely that there was a learning effect on patients as we set a 2-months interval between the two sessions to avoid this bias. This outcome has to be treated cautiously owing to the lost raters in the second session.

Limits of our study

Using videos could be considered as a limit as it is not as faithful as face to face assessment and adds doubts. Indeed, face to face examination offers the opportunity to ask the patient to repeat certain movements and may make the scoring easier.

Given that we are in close relationship with speech language therapists, their over-representation is due to a recruitment bias. This could reduce the power of our study to identify variability factors.
The raters inclusion was based on volunteering. The length of visioning and rating (64.89 min) explains the lost to follow up raters in the second session. It also brings to light the fact that, in clinical practice, the use of SFGS with all the subscores calculation is not practicable in emergency. The comparison with the other studies was difficult as long as the methodologies were all different.

**CONCLUSION**

Based on our results, the SFGS is a reproducible scale from one time to another, showing an important too almost perfect repeatability. It is also reproducible from one rater to another, with almost perfect agreement for the composite score and voluntary movement, and moderate for resting symmetry and synkinesis. The SFGS can be used with good reproducibility by both novices and experts, and by all professionals involved in the management of PFP.

**DISCLOSURE**

The authors report no conflict of interest.
REFERENCES


Table 1: Intra-rater variability for Sunnybrook Facial Grading System composite score and 3 subscores by intraclass correlation coefficient (ICC) and the 95% confidence interval (CI).

<table>
<thead>
<tr>
<th>Raters</th>
<th>Composite score</th>
<th>Resting symmetry</th>
<th>Voluntary movement</th>
<th>Synkinesis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC 95% CI</td>
<td>ICC 95% CI</td>
<td>ICC 95% CI</td>
<td>ICC 95% CI</td>
</tr>
<tr>
<td>Total population (n = 25)</td>
<td>0.915 0.900-0.929</td>
<td>0.694 0.646-0.737</td>
<td>0.903 0.886-0.918</td>
<td>0.810 0.778-0.838</td>
</tr>
<tr>
<td>ENT specialists (n = 5)</td>
<td>0.938 0.909-0.958</td>
<td>0.752 0.653-0.826</td>
<td>0.948 0.924-0.965</td>
<td>0.839 0.770-0.908</td>
</tr>
<tr>
<td>Physiotherapists (n = 3)</td>
<td>0.806 0.667-0.886</td>
<td>0.733 0.591-0.831</td>
<td>0.772 0.623-0.863</td>
<td>0.734 0.593-0.832</td>
</tr>
<tr>
<td>Speech therapists (n = 17)</td>
<td>0.924 0.907-0.938</td>
<td>0.667 0.603-0.722</td>
<td>0.908 0.887-0.925</td>
<td>0.807 0.767-0.841</td>
</tr>
<tr>
<td>Seniors (n = 11)</td>
<td>0.911 0.886-0.931</td>
<td>0.791 0.736-0.836</td>
<td>0.898 0.869-0.921</td>
<td>0.880 0.847-0.907</td>
</tr>
<tr>
<td>Juniors (n = 14)</td>
<td>0.919 0.899-0.935</td>
<td>0.620 0.543-0.687</td>
<td>0.908 0.885-0.927</td>
<td>0.726 0.665-0.777</td>
</tr>
</tbody>
</table>
Table 2: Intra-rater variability and inter-rater variability for Sunnybrook Facial Grading System voluntary movement and synkinesis scores and each component by intraclass correlation coefficient (ICC) and the 95% confidence interval (CI).

<table>
<thead>
<tr>
<th>Subscores</th>
<th>Components</th>
<th>Intra-rater variability</th>
<th>Inter-rater variability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ICC</td>
<td>95% CI</td>
</tr>
<tr>
<td><strong>Voluntary movement</strong></td>
<td>Total</td>
<td>0.903</td>
<td>0.886-0.918</td>
</tr>
<tr>
<td></td>
<td>Forehead wrinkle</td>
<td>0.866</td>
<td>0.843-0.887</td>
</tr>
<tr>
<td></td>
<td>Gentle eye closure</td>
<td>0.817</td>
<td>0.785-0.844</td>
</tr>
<tr>
<td></td>
<td>Open mouth smile</td>
<td>0.874</td>
<td>0.851-0.893</td>
</tr>
<tr>
<td></td>
<td>Snarl</td>
<td>0.844</td>
<td>0.816-0.868</td>
</tr>
<tr>
<td></td>
<td>Lip pucker</td>
<td>0.784</td>
<td>0.747-0.815</td>
</tr>
<tr>
<td><strong>Synkinesis</strong></td>
<td>Total</td>
<td>0.810</td>
<td>0.778-0.838</td>
</tr>
<tr>
<td></td>
<td>Forehead wrinkle</td>
<td>0.661</td>
<td>0.609-0.708</td>
</tr>
<tr>
<td></td>
<td>Gentle eye closure</td>
<td>0.600</td>
<td>0.541-0.653</td>
</tr>
<tr>
<td></td>
<td>Open mouth smile</td>
<td>0.743</td>
<td>0.667-0.754</td>
</tr>
<tr>
<td></td>
<td>Snarl</td>
<td>0.651</td>
<td>0.597-0.698</td>
</tr>
<tr>
<td></td>
<td>Lip pucker</td>
<td>0.761</td>
<td>0.721-0.795</td>
</tr>
</tbody>
</table>
Table 3: Inter-rater variability for Sunnybrook Facial Grading System composite score and 3 subscores by intraclass correlation coefficient (ICC) and the 95% confidence interval (CI).

<table>
<thead>
<tr>
<th>Raters</th>
<th>Composite score</th>
<th>Resting symmetry</th>
<th>Voluntary movement (total)</th>
<th>Synkinesis (total)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC  95% CI</td>
<td>ICC  95% CI</td>
<td>ICC  95% CI</td>
<td>ICC  95% CI</td>
</tr>
<tr>
<td>Total population (n = 31)</td>
<td>0.847 0.755-0.923</td>
<td>0.553 0.408-0.730</td>
<td>0.839 0.746-0.919</td>
<td>0.476 0.467-0.774</td>
</tr>
<tr>
<td>ENT specialists (n = 6)</td>
<td>0.894 0.811-0.951</td>
<td>0.699 0.538-0.842</td>
<td>0.886 0.803-0.946</td>
<td>0.445 0.581-0.862</td>
</tr>
<tr>
<td>Physiotherapists (n = 4)</td>
<td>0.709 0.508-0.858</td>
<td>0.460 0.239-0.691</td>
<td>0.599 0.323-0.802</td>
<td><strong>0.136 0.322-0.750</strong></td>
</tr>
<tr>
<td>Speech therapists (n = 21)</td>
<td>0.885 0.808-0.944</td>
<td>0.573 0.424-0.747</td>
<td>0.880 0.803-0.941</td>
<td>0.568 0.440-0.762</td>
</tr>
<tr>
<td>Seniors (n = 16)</td>
<td>0.849 0.756-0.925</td>
<td>0.546 0.389-0.729</td>
<td>0.836 0.739-0.918</td>
<td>0.484 0.490-0.797</td>
</tr>
<tr>
<td>Juniors (n = 15)</td>
<td>0.848 0.744-0.926</td>
<td>0.545 0.390-0.728</td>
<td>0.851 0.753-0.927</td>
<td>0.469 0.427-0.758</td>
</tr>
</tbody>
</table>