



**HAL**  
open science

## **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:**

Camille Cabrol, Léa Elarouti, Anne-Laure Montava, Sylvie Jarze, Julien Mancini, et al.. Sunnybrook Facial Grading System: Intra-rater and Inter-rater Variabilities. *Otology and Neurotology*, 2021, 42 (7), pp.1089-1094. 10.1097/MAO.0000000000003140 . inserm-03650974

**HAL Id: inserm-03650974**

**<https://inserm.hal.science/inserm-03650974v1>**

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

1 **Sunnybrook Facial Grading System: Intra-rater and inter-rater variabilities**

2

3 **ABSTRACT:**

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

6 *Study design:* Prospective test of hypothesis.

7 *Setting:* University tertiary referral centre.

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

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

17 *Results:* For the total population, the intra-rater ICC was 0.915 [0.900-0.929] for the composite  
18 score considered to represent almost perfect repeatability. Repeatability was important for  
19 symmetry at rest (0.694 [0.646-0.737]), almost perfect for voluntary movements (0.903 [0.886-  
20 0.918]) and synkinesis (0.810 [0.778-0.838]). The inter-rater ICC for the composite score was  
21 0.847 [0.755-0.923] indicating almost perfect agreement between all raters. Agreement  
22 between raters was almost perfect for voluntary movements (0.839 [0.746-0.919]), but  
23 moderate for symmetry at rest (0.553 [0.408-0.730]) and synkinesis (0.476 [0.333-0.666]).  
24 Some differences were found between raters groups; however, repeatability and agreement  
25 were good for all raters.

26 *Conclusions:* The SFGS is a reproducible scale. It can be used with good reproducibility by  
27 both novices and experts, and by all professionals involved in the management of PFP.

28

29 **KEYWORDS:**

30 Facial palsy,

31 Facial grading scales,

32 Sunnybrook Facial Grading System,

33 Inter-rater variability,

34 Intra-rater variability.

35

36

37 **INTRODUCTION**

38

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

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

51 Introduced by Ross in 1996 (2), SFGS assesses facial resting symmetry compared to normal  
52 side, symmetry of voluntary movement and potential synkinesis associated with specified  
53 voluntary movement. Three subscores are obtained giving a composite score from 0 for  
54 complete PFP to 100 for normal facial function. However, the SFGS remains a subjective scale  
55 which makes it subject to limitation because of varying degrees of variability. From 2000 to  
56 2010, some studies on SFGS's variability concluded that SFGS was generally reliable (3–7).  
57 Nevertheless, they found contradictory results of variability for individual scores.

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

61

62 **MATERIALS AND METHODS**

63

64 *Ethical considerations*

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

68

69 *Study protocol*

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

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

81 A group of 31 independent raters was constituted by senior and junior professionals involved  
82 in the management of patients with PFP: ENT specialists, physiotherapists and speech  
83 therapists. The raters analysed the videotapes and graded the facial function of each subject  
84 using the SFGS and the HBGS in 2 independent and timed rounds with a 2-months interval.  
85 Neither documentation nor training session were provided. During session, the raters could  
86 pause or go back at any time if needed to have a better analysis as long as the chronometer was

87 not stopped. The time of evaluation was reported. Moreover, information about their experience  
88 in the management of patients with PFP and a signed consent were reported.

89

### 90 *Statistical analysis*

91 To estimate intra-rater and inter-rater variabilities, the intraclass correlation coefficient (ICC  
92 [95% confidence interval]) was estimated (8). ICC was estimated for the composite score, the  
93 three subscores and each movement for the total population and for each group of raters.

94

95 According to Landis Koch (Biometrics, 1977), we considered the agreement as weak if rated  
96 within 0-0.40, moderate within 0.41-0.60, important within 0.61-0.80 and almost perfect within  
97 0.81-0.99. Analysis were carried out with SPSS 20.0 for Windows.

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

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

104

## 105 **RESULTS**

106

107 In the first session, 31 raters participated: 6 ENT specialists (4 seniors and 2 juniors), 4  
108 physiotherapists (3 seniors and 1 junior) and 21 speech therapists (9 seniors and 12 juniors). In  
109 the second session, 25 of them (80.6%) took part: 5 ENT specialists (3 seniors and 2 juniors),  
110 3 physiotherapists (2 seniors and 1 junior) and 17 speech therapists (6 seniors and 11 juniors).

111 The average time of evaluation was 64.9 minutes (min) for both sessions with a standard  
112 deviation of 22.8. It was  $70.9 \pm 23.4$  min for session 1, and  $57.5 \pm 19.8$  min for session 2.

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

117

### 118 *Intra-rater variability (repeatability)*

119 Results of intra-rater variability for the composite score and the three subscores are reported in  
120 Table 1. The composite score ICC for the total population was 0.915 [0.900-0.929] considered  
121 to represent almost perfect repeatability. Repeatability was important for symmetry at rest  
122 (0.694 [0.646-0.737]), almost perfect for voluntary movement (0.903 [0.886-0.918]) and  
123 synkinesis (0.810 [0.778-0.838]).

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

128 No difference of repeatability was found between our raters groups of seniors and juniors,  
129 except for resting symmetry with better repeatability for seniors (0.791 [0.736-0.836]) than for  
130 juniors (0.620 [0.543-0.687]) and synkinesis with again better repeatability for seniors (0.880  
131 [0.847-0.907]) than for juniors (0.726 [0.665-0.777]).

132 No difference was found between our raters groups of professions, except for the composite  
133 score and voluntary movement with better repeatability for ENT specialists and speech  
134 therapists than for physiotherapists.

135 For extreme HB grades (II and VI), repeatability was almost perfect for the composite score  
136 (0.962 [0.947-0.972]) and the voluntary movement (0.961 [0.946-0.972]), important for resting  
137 symmetry (0.723 [0.636-0.792]) and moderate for synkinesis (0.576 [0.458-0.674]). For  
138 medium HB grades (III, IV, V), repeatability was almost perfect for the composite score (0.836  
139 [0.802-0.865]) and the voluntary movement (0.828 [0.792-0.859]) and important for resting  
140 symmetry (0.677 [0.616-0.730]) and for synkinesis (0.786 [0.742-0.823]). Repeatability was  
141 better for the composite score and voluntary movement of extreme HB grades than medium HB  
142 grades. It was worse for synkinesis and there was no difference for resting symmetry.

143

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

145 Results of inter-rater variability for the composite score and the three subscores are reported in  
146 Table 3. The ICC for the composite score was 0.847 [0.755-0.923] indicating almost perfect  
147 agreement between all raters. Agreement was almost perfect for voluntary movement (0.839  
148 [0.746-0.919]), but moderate for symmetry at rest (0.553 [0.408-0.730]) and synkinesis (0.476  
149 [0.333-0.666]).

150 Results of inter-rater variability for the five standard expressions in voluntary movement and  
151 synkinesis are reported Table 2. For voluntary movement, agreement was almost perfect for  
152 forehead wrinkle (0.805 [0.701-0.899]) and open mouth smile (0.816 [0.714-0.905]). It was  
153 important for gentle eye closure (0.724 [0.588-0.852]), snarl (0.754 [0.629-0.870]) and lip  
154 pucker (0.611 [0.467-0.774]). For synkinesis, agreement was weak for forehead wrinkle (0.387  
155 [0.258-0.582]), gentle eye closure (0.300 [0.188-0.488]), open mouth smile (0.373 [0.246-  
156 0.568]) and snarl (0.340 [0.219-0.532]). It was moderated for lip pucker (0.439 [0.302-0.632]).  
157 No differences between seniors and juniors raters groups were reported. No difference was  
158 found between our raters groups of professions, except for voluntary movement where  
159 physiotherapists had worse agreement than ENT specialists and speech therapists.

160 For extreme HB grades (II and VI), agreement was almost perfect for the composite score  
161 (0.948 [0.873-0.991]) and the voluntary movement (0.954 [0.887-0.992]), important for resting  
162 symmetry (0.738 [0.511-0.945]) and weak for synkinesis (0.018 [-0.004-0.169]). For medium  
163 HB grades (III, IV, V), agreement was important for the composite score (0.691 [0.529-0.855])  
164 and the voluntary movements (0.704 [0.546-0.863]), moderate for resting symmetry (0.461  
165 [0.299-0.696]) and for synkinesis (0.427 [0.269-0.666]). Repeatability of extreme HB grades  
166 was better than medium HB grades for the composite score and voluntary movement, and worse  
167 for synkinesis. There was no difference for resting symmetry.

168

## 169 **DISCUSSION**

170

171 SFGS assesses resting symmetry compared to normal side, symmetry of voluntary movement  
172 and potential synkinesis associated with specified voluntary movement. Resting symmetry is  
173 assessed by a comparison to the normal side of the palpebral fissure (normal, narrow, wide),  
174 the naso-labial fold (normal, absent, less or more pronounced), and the corner of the mouth  
175 (normal, drooped or pulled up/out). The rating is done through a points-giving system (0, 1, 2).  
176 Then, different regions of the face are examined separately, with five standard expressions used  
177 to assess the symmetry of voluntary movement and the degree of synkinesis associated with  
178 movement. The five standard expressions reflect the motor function of the five peripheral  
179 branches of the facial nerve: forehead wrinkle (frontalis), gentle eye closure (orbicularis oculi),  
180 open mouth smile (zygomaticus and risorius), snarl (levator labii superioris alaeque nasi and  
181 levator labii superioris) and lip pucker (orbicularis oris superior and inferior). The symmetry of  
182 voluntary movement for each standard expression is graded on a five-points scale from 1 (no  
183 movement) to 5 (movement complete), depending on the degree of muscle excursion compared  
184 to normal side. The degree of synkinesis associated with each standard expression is rated on a

185 four-points scale from 0 (no synkinesis) to 3 (severe synkinesis). Three scores are obtained and  
186 weighted as follows: the resting symmetry score is multiplied by five, and the voluntary  
187 movement score is multiplied by four. Then, a composite score is calculated by subtracting the  
188 resting symmetry score, and the synkinesis score from the voluntary movement score. A  
189 composite score of 100 corresponds to normal facial function and a composite score of 0 to  
190 complete PFP.

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

199

#### 200 *Intra-rater variability (repeatability)*

201 In our study with 31 raters, repeatability evaluated by ICC for the composite score was 0.900  
202 to 0.929. ICC varied from 0.838 to 0.929 with eight novice assessors in Hu et al. (4), 0.864 to  
203 0.995 with 26 doctors in Kanerva et al. (6), 0.948 to 0.970 for 2 naïve raters in Neely et al. (7).  
204 These results are considered to represent almost perfect repeatability.

205 Although the repeatability score for resting symmetry was considered as important (0.694  
206 [0.646-0.737]) in the current study, it was almost perfect for Kanerva et al. (0.841 [0.500-  
207 0.976]). However, Kanerva et al. specified that results with coefficient of repeatability (CR)  
208 varied for resting symmetry: repeatability was only moderate or fair.

209 We found an almost perfect repeatability for the voluntary movement score (0.903 [0.886-  
210 0.918]) and for each movement performed. This result is consistent with the findings of Kanerva  
211 et al. (0.918 [0.799-0.988]).

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

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

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

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

227

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

229 In our study with 31 raters, agreement evaluated by ICC for the composite score was 0.847  
230 [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  
231 et al. with 2 novice users, 0.997 [0.992-1.000] in Kanerva et al. with 26 doctors, 0.890 [0.784-  
232 0.946] in Neely et al. with 2 naïve raters. These results represent an almost perfect agreement  
233 between raters.

234 We found a moderate agreement for symmetry at rest (0.563 [0.408-0.730]). However, Hu et  
235 al. found an almost perfect agreement (0.950) as Kanerva et al. (0.983 [0.960-0.996]). Kayhan  
236 et al. found an important agreement (0.72 [0.58-0.84]). In the current study, agreement between  
237 raters was almost perfect for voluntary movement (0.839 [0.746-0.919]). This outcome is  
238 congruent with Kayhan et al. (0.83 [0.73-0.90]), Hu et al. (0.976) and Kanerva et al. (0.997  
239 [0.992-0.999]). Coulson et al. (5), with 6 ENT specialists, found important agreement (0.63).  
240 Agreement was almost found perfect for forehead wrinkle and open mouth smile, and important  
241 for gentle eye closure, snarl and lip pucker. Neely et al. specified that forehead wrinkle and lip  
242 pucker were most variable and significantly different from the other movements.  
243 For synkinesis we found moderated agreement (0.476 [0.333-0.666]). It was almost perfect in  
244 Hu et al. (0.913) and Kanerva et al. (0.987 [0.969-0.997]), but they didn't find as good results  
245 using CR to assess agreement: it was, only moderate or fair. Agreement was important for  
246 Kayhan et al. (0.70 [0.55-0.93]). Coulson et al. found a weak agreement (0.23).  
247 In our study, agreement was weak for forehead wrinkle, gentle eye closure, open mouth smile  
248 and snarl, and moderate for lip pucker. These results are opposite to Kayhan et al. who found  
249 an important agreement for each movement except frontal synkinesis which showed a weak  
250 agreement (0.38 [0.21-0.59]). Thus, agreement between raters for synkinesis was highly  
251 variable. Those discrepancies can be due to the fact that the rater has to observe the whole face,  
252 whereas for voluntary movement the rater's attention is directed to a specific region of the face.  
253 It should be noted that in some previous studies (3,4,6,7), raters had preliminary training before  
254 the first session which may have helped to reach higher results of agreement with SFGS than  
255 ours.  
256 Differences between seniors and juniors raters groups were not significant. This result is  
257 consistent with Hu et al. and Kanerva et al. and suggest that the SFGS can be used with little  
258 prior knowledge of the scale and then as reliably by both experts and novice users.

259 No difference was found between our raters groups of professions, except for voluntary  
260 movement with better repeatability for ENT specialists and speech therapists than for  
261 physiotherapists. It suggests that agreement on SFGS is the same for all professionals involved  
262 in the management of PFP, except for physiotherapists. Again, this outcome has to be treated  
263 cautiously owing to the heterogeneous sizes of the raters groups.

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

267

#### 268 *Time of scoring*

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

275

#### 276 *Limits of our study*

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

280 Given that we are in close relationship with speech language therapists, their over-  
281 representation is due to a recruitment bias. This could reduce the power of our study to identify  
282 variability factors.

283 The raters inclusion was based on volunteering. The length of visioning and rating (64.89 min)  
284 explains the lost to follow up raters in the second session. It also brings to light the fact that, in  
285 clinical practice, the use of SFGS with all the subscores calculation is not practicable in  
286 emergency.

287 The comparison with the other studies was difficult as long as the methodologies were all  
288 different.

289

## 290 CONCLUSION

291

292 Based on our results, the SFGS is a reproducible scale from one time to another, showing an  
293 important too almost perfect repeatability. It is also reproducible from one rater to another, with  
294 almost perfect agreement for the composite score and voluntary movement, and moderate for  
295 resting symmetry and synkinesis.

296 The SFGS can be used with good reproducibility by both novices and experts, and by all  
297 professionals involved in the management of PFP.

298

## 299 DISCLOSURE

300 The authors report no conflict of interest.

301

302 **REFERENCES**

303

- 304 1. Fattah AY, Gurusinghe ADR, Gavilan J, et al. Facial Nerve Grading Instruments:  
305 Systematic Review of the Literature and Suggestion for Uniformity. *Plast Reconstr Surg*  
306 2015;135(2):569- 79.
- 307 2. Ross BG, Fradet G, Nedzelski JM. Development of a sensitive clinical facial grading  
308 system. *Otolaryngol-Head Neck Surg* 1996;114:380- 6.
- 309 3. Kayhan FT, Zurakowski D, Rauch SD. Toronto Facial Grading System: Interobserver  
310 reliability. *Otolaryngol-Head Neck Surg* 2000;122(2):212- 5.
- 311 4. Hu W, Ross B, Nedzelski J. Reliability of the Sunnybrook Facial Grading System by novice  
312 users. *J Otolaryngol* 2001;30(4).
- 313 5. Coulson SE, Croxson GR, Adams RD, O'dwyer NJ. Reliability of the "Sydney,"  
314 "Sunnybrook," and "House Brackmann" Facial Grading Systems to Assess Voluntary  
315 Movement and Synkinesis after Facial Nerve Paralysis. *Otolaryngol-Head Neck Surg*  
316 2005;132(4):543- 9.
- 317 6. Kanerva M, Poussa T, Pitkäranta A. Sunnybrook and House-Brackmann Facial Grading  
318 Systems: Intrarater repeatability and interrater agreement. *Otolaryngol-Head Neck Surg*  
319 2006;135(6):865- 71.
- 320 7. Neely JG, Cherian NG, Dickerson CB, Nedzelski JM. Sunnybrook facial grading system:  
321 Reliability and criteria for grading. *The Laryngoscope* 2010;120:1038- 45.
- 322 8. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status  
323 measures. Statistics and strategies for evaluation. *Control Clin Trials* 1991;12(4  
324 Suppl):142S-158S.

325

326

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).

Raters	Composite score		Resting symmetry		Voluntary movement		Synkinesis	
	ICC	95% CI	ICC	95% CI	ICC	95% CI	ICC	95% CI
<b>Total population (n = 25)</b>	0.915	0.900-0.929	0.694	0.646-0.737	0.903	0.886-0.918	0.810	0.778-0.838
ENT specialists (n = 5)	0.938	0.909-0.958	0.752	0.653-0.826	0.948	0.924-0.965	0.839	0.770-0.908
Physiotherapists (n = 3)	0.806	0.667-0.886	0.733	0.591-0.831	0.772	0.623-0.863	0.734	0.593-0.832
Speech therapists (n = 17)	0.924	0.907-0.938	0.667	0.603-0.722	0.908	0.887-0.925	0.807	0.767-0.841
Seniors (n = 11)	0.911	0.886-0.931	0.791	0.736-0.836	0.898	0.869-0.921	0.880	0.847-0.907
Juniors (n = 14)	0.919	0.899-0.935	0.620	0.543-0.687	0.908	0.885-0.927	0.726	0.665-0.777

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).

Subscores	Components	Intra-rater variability		Inter-rater variability	
		ICC	95% CI	ICC	95% CI
<b>Voluntary movement</b>	<b>Total</b>	0.903	0.886-0.918	0.839	0.746-0.919
	Forehead wrinkle	0.866	0.843-0.887	0.805	0.701-0.899
	Gentle eye closure	0.817	0.785-0.844	0.724	0.588-0.852
	Open mouth smile	0.874	0.851-0.893	0.816	0.714-0.905
	Snarl	0.844	0.816-0.868	0.754	0.629-0.870
	Lip pucker	0.784	0.747-0.815	0.611	0.467-0.774
<b>Synkinesis</b>	<b>Total</b>	0.810	0.778-0.838	0.476	0.333-0.666
	Forehead wrinkle	0.661	0.609-0.708	0.387	0.258-0.582
	Gentle eye closure	0.600	0.541-0.653	0.300	0.188-0.488
	Open mouth smile	0.743	0.667-0.754	0.373	0.246-0.568
	Snarl	0.651	0.597-0.698	0.340	0.219-0.532
	Lip pucker	0.761	0.721-0.795	0.439	0.302-0.632

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).

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