

*Automated objective and marker-free facial grading using photographs of patients with facial palsy*

**Oliver Mothes, Luise Modersohn, Gerd Fabian Volk, Carsten Klingner, Otto W. Witte, Peter Schlattmann, Joachim Denzler, et al.**

**European Archives of Oto-Rhino-Laryngology**  
and Head & Neck

ISSN 0937-4477

Eur Arch Otorhinolaryngol  
DOI 10.1007/s00405-019-05647-7



**Your article is protected by copyright and all rights are held exclusively by Springer-Verlag GmbH Germany, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**



# Automated objective and marker-free facial grading using photographs of patients with facial palsy

Oliver Mothes<sup>1</sup> · Luise Modersohn<sup>1</sup> · Gerd Fabian Volk<sup>2,3</sup> · Carsten Klingner<sup>3,4</sup> · Otto W. Witte<sup>3,4</sup> · Peter Schlattmann<sup>5</sup> · Joachim Denzler<sup>1</sup> · Orlando Guntinas-Lichius<sup>2,3</sup>

Received: 26 June 2019 / Accepted: 11 September 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

**Purpose** An automated, objective, fast and simple classification system for the grading of facial palsy (FP) is lacking.

**Methods** An observational single center study was performed. 4572 photographs of 233 patients with unilateral peripheral FP were subjectively rated and automatically analyzed applying a machine learning approach including Supervised Descent Method. This allowed an automated grading of all photographs according to House-Brackmann grading scale (HB), Sunnybrook grading system (SB), and Stennert index (SI).

**Results** Median time to first assessment was 6 days after onset. At first examination, the median objective HB, total SB, and total SI were grade 3, 45, and 5, respectively. The best correlation between subjective and objective grading was seen for SB and SI movement score ( $r=0.746$ ;  $r=0.732$ , respectively). No agreement was found between subjective and objective HB grading [Test for symmetry 80.61,  $df = 15$ ,  $p < 0.001$ , weighted kappa =  $-0.0105$ ; 95% confidence interval (CI) =  $-0.0542$  to  $0.0331$ ;  $p = 0.6541$ ]. Also no agreement was found between subjective and objective total SI (test for symmetry 166.37,  $df = 55$ ,  $p < 0.001$ ) although there was a nonzero weighted kappa =  $0.2670$ ; CI  $0.2154-0.3186$ ;  $p < 0.0001$ ). Based on a multinomial logistic regression the probability for higher scores was higher for subjective compared to objective SI (OR 1.608; CI 1.202–2.150;  $p = 0.0014$ ). The best agreement was seen between subjective and objective SB (ICC =  $0.34645$ ).

**Conclusions** Automated Sunnybrook grading delivered with fair agreement fast and objective global and regional data on facial motor function for use in clinical routine and clinical trials.

**Keywords** Facial image analysis · Facial paralysis · Active appearance model · Random decision forest · Diagnostics · Grading

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00405-019-05647-7>) contains supplementary material, which is available to authorized users.

✉ Orlando Guntinas-Lichius  
orlando.guntinas@med.uni-jena.de

<sup>1</sup> Department of Computer Science, Friedrich Schiller University Jena, Jena, Germany

<sup>2</sup> Department of Otorhinolaryngology, Jena University Hospital, Am Klinikum 1, 07747 Jena, Germany

<sup>3</sup> Facial Nerve Center, Jena University Hospital, Jena, Germany

<sup>4</sup> Department of Neurology, Jena University Hospital, Jena, Germany

<sup>5</sup> Department of Medical Statistics, Computer Sciences and Data Science, Jena University Hospital, Jena, Germany

## Introduction

Facial palsy (FP) not only leads to a variety of motor deficits in the face, but also affects the emotional and expressive possibilities of the patient. The psychosocial effects have received more attention in the recent years [1, 2]. To assess the psychosocial impairment internationally accepted general psychological measurement tools and also patient-reported outcomes measures like the FDI and FaCE have been established [2–4]. In contrast, only little progress has been made toward the development of an internationally accepted facial nerve grading scale to assess the clinically dominant motor disorder related to facial palsy [5, 6]. The main reason probably is that all facial gradings systems used in clinical routine rely on a subjective assessment. Due to the subjective nature of the assessment, all grading systems, even the frequently used systems like the House-Brackmann

and Sunnybrook scale, have problems with reproducibility with low interobserver and intraobserver variability, and sensitivity to changes over time and/or following interventions [5]. One should be aware that new electronic scales are practical because they can be used in a variety of electronic devices, including smartphones, tablets, and computers [7]. Nevertheless, such systems still are clinician based and not objective. Although these deficits are well known, even the very important randomized controlled clinical trials on treatment of FP relied and still rely on these subjective assessment tools [8–10].

There is an urgent need for photographic standardization to report outcomes for patients with facial palsy [11]. Furthermore, a reliable grading tool offering zonal information of the face, describing static, dynamic, and synkinesis features extracted from these photographic series is needed [12]. At best, the results should be presented to the clinician with immediate graphical outputs [13].

It is said that quantitative technology-based systems still are too complex for clinical routine, are not yet widely available and are expensive [5]. Recently, we have introduced an easy to use, fast, automated and objective facial grading system based on action coding of facial expressions using standardized photographs [14]. Facial action coding is a standard method to assess facial expression in psychology but not familiar to most physicians dealing with facial palsy. Now we refined the system allowing a presentation of the objective results with graphical outputs in the notation of the Stennert index, House-Brackmann scale, and Sunnybrook grading, three facial grading systems physicians are much more familiar with for application in clinical routine treatment of patients with facial palsy.

## Material and methods

The study protocol for this retrospective data analysis was approved by the institutional ethics committee.

### Patients

A standardized data collection were performed in the Department of blinded. All photograph series of patients with unilateral FP were collected from January 2007 to December 2011. The diagnostic procedures and therapy for all patients were the same, for details, see [15].

### Standard photographs of facial expression

Patients with acute facial palsy are routinely photographed using static posed facial expressions in a standardized manner. Participants were seated in a brightly lit room and instructed to focus a digital camera. The patients were

instructed to perform a facial expression as demonstrated by the professional photographer. The patients were requested to perform the expression spontaneously. Furthermore, facial expressions were not exercised prior to the photographs. The sequence of static posed nine expressions was always constant (Supplementary material: Supplement Fig. 1): (1) at rest, (2) closing both eyes, (3) closing both eyes with maximal effort, (4) frowning, (5) wrinkling the nose, (6) lifting corners of mouth with closed mouth, (7) showing the teeth, (8) pursing the lips, and (9) pull down both corners of mouths. Hence, normally nine images were taken as a set per patient per time of assessment. All images were transferred serially from the digital camera into the electronic medical record of patients.

### Clinical grading of facial palsy

The FP was graded according to the House-Brackmann grading [16], the Sunnybrook grading [17] and also according to the Stennert Index [18]. The House-Brackmann scale is a gross six-point facial grading system ( $I$ =normal;  $VI$ =total paralysis). Sunnybrook grading is a regional weighted system that rates three subscores: resting symmetry, the degree of voluntary facial muscle movement, involuntary muscle contraction (synkinesis). The three subscores are used to calculate a composite score (0=total paralysis; 100=normal function). The Stennert index classifies the face at rest (0–4 points; 0=normal to 4=complete loss of resting tone) and during motion (0–6 points; 0=normal to 6=no motion) separately. Both subscores are summed up to the Stennert total score. Two trained observers independently graded all photograph series. In case of disagreement, the photograph series was evaluated together to find an agreement.

### Automated facial grading using compact discriminative facial image features and a support vector machine

A novel fast and marker-free automated method for unilateral facial grading was developed based on support vector machines [19] and certain discriminative facial image features [20–23]. Previously, we have published an active appearance model (AAM) approach for automated action coding of facial expressions in patients with FP [14]. The present method is a further development. Instead of only using features of a trained AAM and Action Units (AU), predicted from it, as described in another publication by our group [21], we additionally exploited Euclidean landmark distances, predicted by Supervised Descent Method (SDM) [22], and features in the form of vectorized layer activations of a pre-trained Convolutional Neural Network (CNN) [23]. Briefly, the contribution is twofold: first, compared to landmark detection by AAM fitting,

the application of SDM has the advantage of numerical stability during localization, improved speed as well as less sensitivity against differences between the expressions used to train the model (usually standard emotions using publicly available datasets) and analyzed expression under facial paralysis that differs a lot from normal emotions. In addition to the AAM model parameter, which covers linear relationships between expressions, and AUs, the distance between landmarks localized by SDM from the left and right side of the face were computed. Moreover, features were extracted out of layer activations of a pre-trained CNN, which show advances in many machine learning applications. The second improvement arises from a different set of features used for the facial grading. Based on different sub-indices in the clinical grading of facial palsy further investigation has shown that particular feature types (of the first contribution) for specific sub-indices are better suited for the prediction process and achieved better results. Support vector machines (SVM) were trained for each sub-index using the related features generated from the training images and ground truth grading (i.e., the subjective grading results) for facial grading prediction.

In summary, methods for generating new, powerful features partially using a more stable localization method of facial landmarks, that showed significant improvement in localization accuracy in the context of faces with FP, were developed. Moreover, the set of new features were used for facial grading prediction.

## Statistical analysis

All statistical analyses were performed using IBM SPSS, version 24.0.0.0 and SAS 9.4 procedures proc mixed and proc freq. Univariate analysis with Spearman's correlation was used to analyze the correlation between subjective and objective gradings. Kappa statistics were used for the ordinal grading of House-Brackmann grading and Stennert index. Furthermore multinomial logistic regression random effects models were used to examine the association between subjective and objective gradings for these scales. The intraclass coefficient (ICC) based on a linear one way random effects model was applied for the scaled data of the Sunnybrook grading. All tests were two-tailed, and  $p$  values  $<0.05$  were considered significant.

## Results

### Patients' characteristics and standard photograph series

Table 1 gives an overview of the patients' characteristics. Median age of the patients was 50 years with a wide range from young children (4 years) to elderly age (91 years). The genders were balanced (Females: 52%). There was no side dominance (Left side: 55%). Most patients had an idiopathic FP (59%), followed by traumatic lesions (19%), and FP due to infectious diseases (16%).

**Table 1** Characteristics of patients with peripheral facial palsy ( $N=233$ )

Parameter	$N$	%
Gender		
Female	122	52.4
Male	111	47.6
Affected side		
Right	104	44.6
Left	129	55.4
Both	0	0
Etiology		
Idiopathic	138	59.2
Neoplastic	6	2.6
Infectious	38	16.3
Traumatic	45	19.3
Miscellaneous	6	2.6
Central	0	0
	Mean $\pm$ SD	Medium, range
Age, years	50.1 $\pm$ 20.5	50, 4–91

SD standard deviation

Overall, 508 standard photograph series, i.e. 4572 still photographs, from 233 patients were analyzed and facial gradings were performed (Supplementary material: Supplemental Table 1). The median interval between the onset of the palsy and first photography series was 6 days. 95% of the patients presented within 90 days after onset. 44.6%, 26.6% and 14.2% of the patients were photographed one time, two times, and three times, respectively, during the course of FP. The median interval between first and second photography session was 41 days. The median interval between second and third photography session was 63 days.

### Subjective and objective facial grading

All subjective and objective gradings for the three gradings system are summarized in Table 2. At first examination, the median subjective and objective House-Brackmann grading were 3 for both assessments (range 1–6 for subjective grading and 2–4 for objective grading). At first examination, the median subjective and objective total scores of Stennert index were 5 for both assessments (range 0–10 for subjective grading and 0–8 for objective grading). At first examination, the median subjective and objective total scores of the Sunnybrook grading were 41 and 45, respectively (range 1–96 for subjective grading and 11–91 for objective grading).

Spearman's correlation analysis is shown in Table 3 and Fig. 1. The correlation in-between the three different grading

**Table 2** Subjective and objective grading of the patients at first examination (N=233)

Parameter	First examination		All examinations	
	Subjective	Objective	Subjective	Objective
<b>House-Brackmann grading</b>				
House-Brackmann grading	3.3 ± 1.1	3.1 ± 0.3	3.3 ± 1.1	3.1 ± 0.3
<b>Sunnybrook grading</b>				
Resting symmetry score	10.8 ± 5.3	9.1 ± 5.2	9.2 ± 5.8	8.8 ± 5.3
Voluntary movement score	59.3 ± 19.5	58.6 ± 14.0	66.3 ± 20.2	64.4 ± 15.6
Synkinesis score	2.6 ± 2.6	2.1 ± 0.9	2.4 ± 2.5	2.0 ± 0.9
Composite score	45.9 ± 24.2	47.5 ± 16.4	54.7 ± 25.6	53.5 ± 18.3
<b>Stennert index</b>				
Index at rest	1.7 ± 1.4	0.9 ± 0.8	1.4 ± 1.3	1.0 ± 0.7
Index during motion	3.6 ± 2.2	4.5 ± 1.3	3.1 ± 2.2	4.3 ± 1.2
Total index	5.3 ± 3.2	5.4 ± 1.6	4.4 ± 3.1	5.2 ± 1.6

SD standard deviation, NA not applicable

Values are mean ± SD

**Table 3** Correlation between the different subjective and objective gradings\*

	Subjective			Objective		
	House-Brackmann	Sunnybrook grading	Stennert index	House-Brackmann	Sunnybrook grading	Stennert index
<b>Subjective</b>						
House-Brackmann	–	$r = -0.512$ $p < 0.0001$	$r = 0.529$ $p < 0.0001$	$r = 0.007$ $p = 0.932$	–	–
Sunnybrook grading	$r = -0.512$ $p < 0.0001$	–	$r = -0.856$ $p < 0.0001$	–	$r = 0.698$ $p < 0.0001$	–
Stennert index	$r = 0.529$ $p < 0.0001$	$r = -0.856$ $p < 0.0001$	–	–	–	$r = 0.534$ $p < 0.0001$
<b>Objective</b>						
House-Brackmann	$r = 0.007$ $p = 0.932$	–	–	–	$r = -0.091$ $p = 0.383$	$r = -0.238$ $p = 0.003$
Sunnybrook grading	–	$r = 0.698$ $p < 0.0001$	–	$r = -0.091$ $p = 0.383$	–	$r = -0.446$ $p < 0.0001$
Stennert index	–	–	$r = 0.534$ $p < 0.0001$	$r = -0.238$ $p = 0.003$	$r = -0.446$ $p < 0.0001$	–

House-Brackmann scale and Stennert index; higher value: worse function. Sunnybrook grading; higher value: better function

systems was moderate ( $r \leq 0.5$ ), both in-between the subjective and objective gradings. There was no correlation between the subjective and objective House-Brackmann grading ( $r = 0.007$ ). The correlation between the subjective and objective total score of the Stennert index was moderate ( $r = 0.534$ ). The correlation was best for the Sunnybrook grading. The correlation between the subjective and objective total score of the Sunnybrook grading was good ( $r = 0.698$ ). An example of a series measurement over the time course of the facial palsy is shown for one patient in Fig. 2.

Kappa and ICC statistics and the multinomial logistic regressions confirmed the results of the univariate correlation analysis. There was no agreement between subjective and objective House-Brackmann grading (Test for symmetry 80.61,  $df = 15$ ,  $p < 0.001$ ) the corresponding weighted kappa =  $-0.0105$ ; 95% confidence interval (CI) =  $-0.0542$  to  $0.0331$ ;  $p = 0.6541$ ). Based on the multinomial logistic regression analysis, there was a lower probability for higher (more severe FP) subjective compared to objective House-Brackmann gradings [Odds ratio (OR) 0.573; CI 0.360–0.912;  $p = 0.0193$ ]. There is also no agreement between subjective and objective total Stennert grading (Test for symmetry 166.37,  $df = 55$ ,  $p < 0.001$ ) although there is nonzero weighted kappa =  $-0.2670$ ; CI 0.2154–0.3186;  $p < 0.0001$ ). Here, the probability for higher scores (more severe FP) was higher for subjective compared to objective Stennert gradings (OR 1.608; CI 1.202–2.150;  $p = 0.0014$ ). The best agreement was seen between subjective and objective Sunnybrook grading (ICC 0.34645;  $p = 0.5596$ ; Fig. 3).

## Discussion

It is suggested that the ideal facial nerve grading instrument should have to follow characteristics [5]: it provides regional scoring of facial function, performs static and dynamic measures, examines secondary sequelae of facial palsy like synkinesis, yields reproducible results with low inter-observer and intraobserver variability, is sensitive to track changes over time and following interventions, convenient for clinical use. In absence of a sophisticated technology-based methodology for clinical application in daily routine, members of the Sir Charles Bell Society have recommended a widespread adoption of the clinician-based Sunnybrook Facial Grading Scale as the current standard in reporting outcomes of facial nerve disorders [24]. As the objective Sunnybrook grading revealed the best results, we primarily propose to use the presented automated algorithm to generate a Sunnybrook grading.

The presented approach seems to fulfill many aspects of an ideal facial grading instrument [11]. It uses nine easy to take standardized photographs. Taking standardized

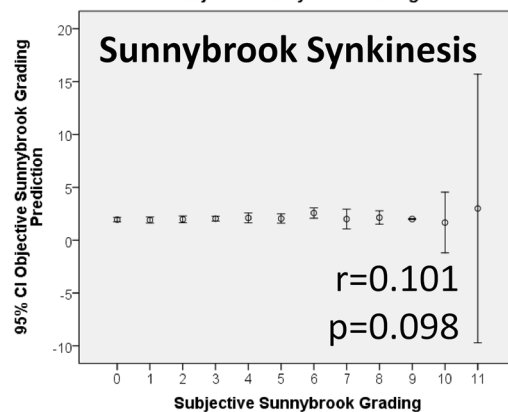
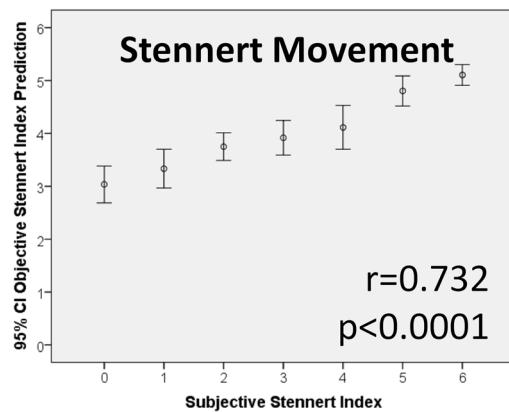
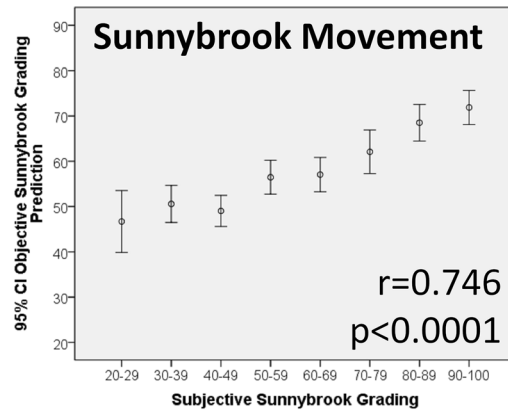
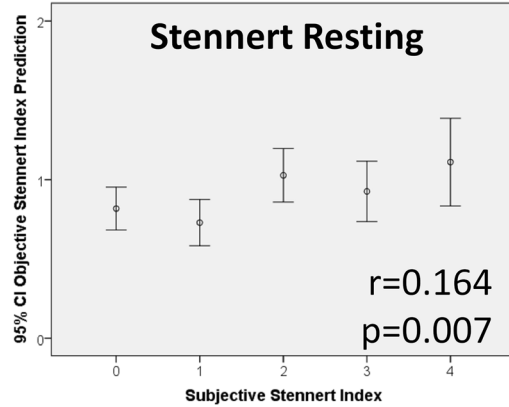
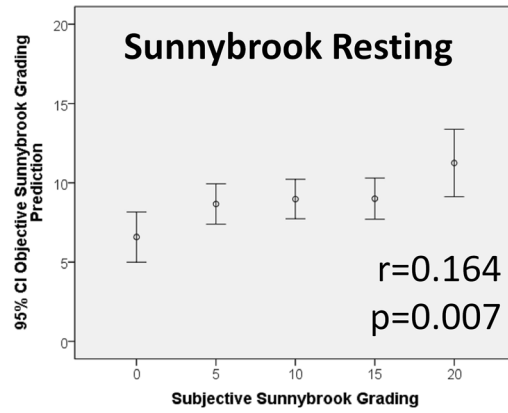
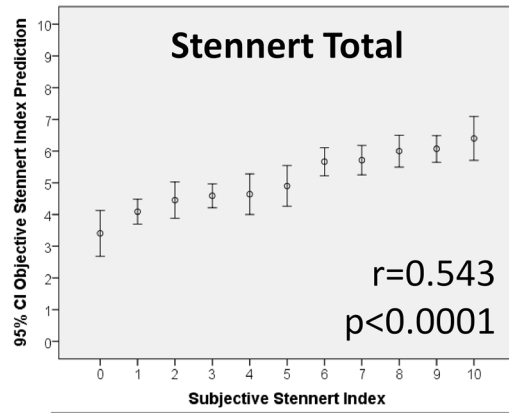
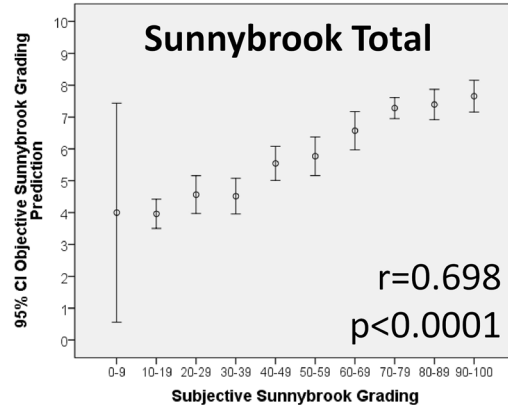
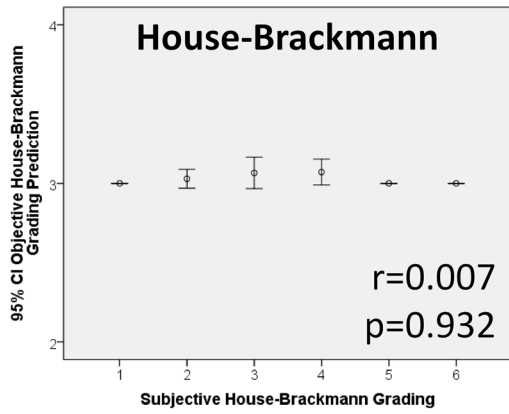
photographs still is much easier to realize in clinical routine than taking a standardized video. The approach is marker-free, i.e. this source of error and variability is excluded. It does not matter how often the photographs are presented to the algorithm: An analysis of the same nine pictures always reveals the same result, i.e. there is absolutely no interobserver and intraobserver variability. Furthermore, the results of the image analysis are expressed in the familiar nomenclature of classical grading system instead of introducing new and unfamiliar parameters to characterize facial function. This was a major disadvantage of our first approach automatically recognizing action units defined by the Facial Action Coding System (FACS) [14]. Psychologists are familiar with FACS, but neurologists and otorhinolaryngologists are not trained to use FACS to define a motor dysfunction in patients with FP.

The Facial Assessment by Computer Evaluation software (FACEgram) is a validated free software developed for accurate measurement of facial dimensions, proportions, and angles based on the analysis of video pixel data using artificial neural networks [25]. It also does not need head markers. The result is expressed as an objective House-Brackmann grading and as relative function in different facial region compared to the healthy side. It was proposed that FACEgram might have the potential to become the universal automated facial grading system [26]. Interestingly, the groups propagating FACEgram actually focus more on spreading an electronic version of a subjective clinician-graded facial function scale [6]. The reasons might be that the FACEgram analysis is limited to patients with normal pupil position and cannot quantify static procedures.

The OSCAR system is also a marker-free system, but needs the recording of a 20 min video to calculate an objective synkinesis index [27]. Although the system was first introduced in 1994, it took until 2017 to present further 13 patients. A relative synkinesis index is generated expressing the amount of asymmetry induced by synkinesis on the paretic face in relation to the active movements of the normal side. This is a quite unfamiliar parameter. Furthermore, the functional deficit induced by the palsy is not quantified.

The development of a new facial grading system is facing the immanent problem that there is no gold standard because all subjective systems including the three systems applied in the present study have a more or less limited accuracy. Putting this aside, the objective Sunnybrook grading revealed the best results. This simultaneously provides the benefit that an objective regional scoring of facial function for static and dynamic measures is calculated, that synkinesis is scored if present, and that a nomenclature is used that is anyway suggested for standard reporting of facial motor function in facial nerve disorders.

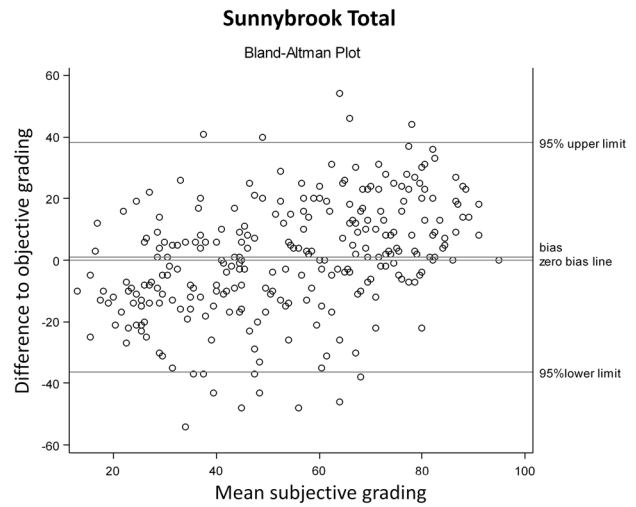
The use of still photographs has immanent limitations as the dysfunction of facial movement is deduced from





**Fig. 1** Comparison of the subjective versus the objective grading by the three different grading systems: House-Brackmann grades (I–VI), Stennert total index (0–10), Stennert index subscores, Sunnybrook total score grading results (0–100), and Sunnybrook subscore gradings. The Sunnybrook grading showed the best prediction and correlation. X-axis: subjective grading. Y-axis: 95% confidence intervals (CI) of predication of the objective grading). The results of Spearman's correlation analysis are also shown

stills. Changes over time during a facial task are not taken into account. Especially the analysis of synkinesis is at best an analysis of synchronous movement over the time. This explains the limited results when looking only on the analyses of synkinetic movements in our patients' cohort. Furthermore, it might be helpful to train the algorithm in a large set of patients with more variable synkinesis. The future certainly lies in a three-dimensional (3D) video analysis of the face. A 3D analysis allows a much more detailed analysis of slightest movements in the face. 3D cameras become cheaper and first marker-based approaches have been published [28].



**Fig. 3** Bland–Altman plot showing the agreement between subjective and objective Sunnybrook grading. A fair agreement can only be seen between subjective and objective Sunnybrook grading



**Fig. 2** Example of a 77-year-old woman with idiopathic facial palsy. Standard photographs, each row is showing one examination. The patients showed an incomplete recovery: good recovery of the resting tone with full symmetry, but defective healing of the muscle move-

ment coordination. The graph is showing the subjective (blue) and objective grading of the photographic series over the time course of the facial nerve regeneration. First examination 3 days after onset of the palsy on top, last examination 696 days after onset at the bottom

Currently, we try to adapt the presented algorithm to analyze 3D videos of patients with facial palsy. Due to the large data volumes generated by 3D videos, the analyses will be, at least for the next years, limited to practices and hospital with appropriate hardware. To track patients at home simpler approaches are needed. A solution would be a smartphone-based automatic diagnosis system. A first approach was published that discriminates facial nerve palsy from normal subjects has been published [29]. These systems have to be expanded to allow regional analyses and validated in larger cohorts of patients.

**Acknowledgement** The authors thank Saskia Schenk for her support in collecting the study data.

**Author contributions** The author who carried out the biostatistical analysis: Peter Schlattmann, Orlando Guntinas-Lichius.

**Funding** This work was supported by the German Federal Ministry of Education and Research (BMBF; Project IRESTRA Grant No.16SV7209) and by the Deutsche Forschungsgemeinschaft (DFG; grant No. GU 463/12-1 and DE 735/15-1).

## Compliance with ethical standards

**Conflict of interest** All authors declare that they do not have a conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## References

- Dobel C, Miltner WH, Witte OW, Volk GF, Guntinas-Lichius O (2013) Emotional impact of facial palsy. *Laryngorhinootologie* 92(1):9–23. <https://doi.org/10.1055/s-0032-1327624>
- Dobel C, Guntinas-Lichius O (2016) Psychological exploration of emotional, communicative, and social impairments in patients with facial impairments. In: Guntinas-Lichius O, Schaitkin BM (eds.) *Facial nerve disorders and diseases: diagnosis and management*. Thieme, Stuttgart
- VanSwearingen JM, Brach JS (1996) The Facial Disability Index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. *Phys Ther* 76(12), 1288–98 (discussion 98–300. Epub 1996/12/01).
- Kahn JB, Gliklich RE, Boyev KP, Stewart MG, Metson RB, McKenna MJ (2001) Validation of a patient-graded instrument for facial nerve paralysis: the FaCE scale. *Laryngoscope* 111(3):387–398. <https://doi.org/10.1097/00005537-200103000-00005> (Epub 2001/02/27)
- Fattah A, Gurusinge A, Gavilan J, Hadlock T, Marcus J, Marres H et al (2014) Facial nerve grading instruments: systematic review of the literature and suggestion for uniformity. *Plast Reconstr Surg*. <https://doi.org/10.1097/PRS.0000000000000905>
- Banks CA, Jowett N, Azizzadeh B, Beurskens C, Bhama P, Borschel G et al (2017) Worldwide testing of the eFACE facial nerve clinician-graded scale. *Plast Reconstr Surg* 139(2):491e–e498. <https://doi.org/10.1097/PRS.0000000000002954>
- Gaudin RA, Robinson M, Banks CA, Baiungo J, Jowett N, Hadlock TA (2016) Emerging vs time-tested methods of facial grading among patients with facial paralysis. *JAMA Facial Plast Surg*. <https://doi.org/10.1001/jamafacial.2016.0025>
- Sullivan FM, Swan IR, Donnan PT, Morrison JM, Smith BH, McKinstry B et al (2007) Early treatment with prednisolone or acyclovir in Bell's palsy. *N Engl J Med* 357(16):1598–1607
- Engstrom M, Berg T, Stjernquist-Desatnik A, Axelsson S, Pitkaranta A, Hultcrantz M et al (2008) Prednisolone and valaciclovir in Bell's palsy: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet Neurol* 7(11):993–1000
- Babl FE, Mackay MT, Borland ML, Herd DW, Kochar A, Hort J et al (2017) Bell's Palsy in Children (BellPIC): protocol for a multicentre, placebo-controlled randomized trial. *BMC Pediatr* 17(1):53. <https://doi.org/10.1186/s12887-016-0702-y>
- Santosa KB, Fattah A, Gavilan J, Hadlock TA, Snyder-Warwick AK (2017) Photographic standards for patients with facial palsy and recommendations by members of the sir charles bell society. *JAMA Facial Plast Surg*. <https://doi.org/10.1001/jamafacial.2016.1883>
- Hadlock TA, Urban LS (2012) Toward a universal, automated facial measurement tool in facial reanimation. *Arch Facial Plast Surg* 14(4):277–282. <https://doi.org/10.1001/archfacial.2012.111>
- Hadlock T (2016) Standard outcome measures in facial paralysis: getting on the same page. *JAMA Facial Plast Surg* 18(2):85–6. <https://doi.org/10.1001/jamafacial.2015.2095> (Epub 2016/01/10)
- Haase D, Minnigerode L, Volk GF, Denzler J, Guntinas-Lichius O (2015) Automated and objective action coding of facial expressions in patients with acute facial palsy. *Eur Arch Oto-rhino-laryngol* 272(5):1259–1267. <https://doi.org/10.1007/s00405-014-3385-8>
- Volk GF, Klingner C, Finkensieper M, Witte OW, Guntinas-Lichius O (2013) Prognostication of recovery time after acute peripheral facial palsy: a prospective cohort study. *BMJ Open*. <https://doi.org/10.1136/bmjopen-2013-003007>
- House JW, Brackmann DE (1985) Facial nerve grading system. *Otolaryngol Head Neck Surg* 93(2):146–7
- Ross BG, Fradet G, Nedzelski JM (1996) Development of a sensitive clinical facial grading system. *Otolaryngol Head Neck Surg*. 114(3):380–6
- Stennert E, Limberg CH, Frentrup KP (1977) An index for paresis and defective healing—an easily applied method for objectively determining therapeutic results in facial paresis (author's transl). *Hno* 25(7):238–245
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–295
- Cootes TF, Edwards CA (2001) Active appearance models. *IEEE Trans Pattern Anal Mach Intell* 23(6):681–685
- Modersohn L, Denzler J (eds.) (2016) Facial paresis index prediction by exploiting active appearance models for compact discriminative features. In: *Proceedings of the 11th joint conference on computer vision, imaging and computer graphics theory and applications (VISIGRAPP 2016)* 2016.
- Xiong X, De la Torre F (2013) Supervised descent method and its applications to face alignment. *IEEE Conf Comput Vis Pattern Recognit*. <https://doi.org/10.1109/CVPR.2013.75>
- Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T (2014) DeCAF: a deep convolutional activation feature for generic visual recognition. In: *Proceedings of the 31st international conference on machine learning 2014; 32 (ICML'14)*, Eric P. Xing and Tony Jebara (Eds.), Vol. 32. *JMLR.org* I-647-I-655
- Fattah AY, Gavilan J, Hadlock TA, Marcus JR, Marres H, Nduka C et al (2014) Survey of methods of facial palsy documentation

- in use by members of the Sir Charles Bell Society. *Laryngoscope* 124(10):2247–51. <https://doi.org/10.1002/lary.24636>
25. O'Reilly BF, Soraghan JJ, McGrenary S, He S (2010) Objective method of assessing and presenting the House-Brackmann and regional grades of facial palsy by production of a facogram. *Otol Neurotol* 31(3):486–91. <https://doi.org/10.1097/MAO.0b013e3181c993dc> (**Epub 2010/01/01**)
26. Lee LN, Susarla SM, Hohman MH, Henstrom DK, Cheney ML, Hadlock TA (2013) A comparison of facial nerve grading systems. *Ann Plast Surg* 70(3):313–6. <https://doi.org/10.1097/SAP.0b013e31826acb2c>
27. Meier-Gallati V, Scriba H (2017) Objective assessment of the reliability of the House-Brackmann and Fisch grading of synkinesis. *Eur Arch Oto-rhino-laryngology* 274(12):4217–4223. <https://doi.org/10.1007/s00405-017-4770-x>
28. Horta R, Nascimento R, Geros A, Aguiar P, Silva A, Amarante J (2018) A novel system for assessing facial muscle movements: the facegram 3D. *Surg Innov* 25(1):90–92. <https://doi.org/10.1177/1553350617753227>
29. Kim HS, Kim SY, Kim YH, Park KS (2015) A smartphone-based automatic diagnosis system for facial nerve palsy. *Sensors (Basel)* 15(10):26756–26768. <https://doi.org/10.3390/s151026756>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.