OPEN

PRS GLOBAL OPEN

ORIGINAL ARTICLE
Cosmetic

Novel Validated Five-point Photonumeric Scales for Assessment of Static and Dynamic Forehead Lines

Tingsong Lim, MD*
Martina Kerscher, MD†
Alexandra Ogilvie, MD‡
Tatjana Pavicic, MD§
Paul Z. Lorenc, MD¶
Konstantin Frank, MD||
Valentina Prinz, MBA**
Mia Cajkovsky**
Rainer Pooth, MD, PhD††
Claudia Hernandez, MD‡‡

Background: The objective of this investigation was to create and validate five-point photonumeric scales which assess static and dynamic forehead lines.
Methods: Two different novel five-point photonumeric scales for the assessment of static and dynamic forehead lines were developed. Moreover, a photoguide was created, including subjects from both sexes, all age groups, and different Fitzpatrick skin types. A total of 11 raters from all over the world were involved in the digital validation, whereas four raters performed a live validation.
Results: The Croma Static Forehead Lines–Assessment Scale showed almost perfect inter and intra-rater agreement in both the digital and the live setting with inter-rater intraclass correlation coefficients of 0.86 [95% confidence interval (CI): 0.82–0.89] in the first digital rating and 0.82 [95% CI: 0.78–0.86] in the second digital rating. The Croma Dynamic Forehead Lines–Assessment Scale showed almost perfect inter and intra-rater agreement in the digital setting with inter-rater intraclass correlation coefficients of 0.83 [95% CI: 0.79–0.86] in the first digital rating and 0.80 [95% CI: 0.75–0.84] in the second rating and almost substantial agreement in the live setting.
Conclusions: The Croma Static Forehead Lines–Assessment Scale and the Croma Dynamic Forehead Lines–Assessment Scale have excellent inter and intra-rater agreements to be justifiably used in the clinical and study setting, both digitally and live across ethnic groups. *(Plast Reconstr Surg Glob Open 2023; 11:e5287; doi: 10.1097/GOX.0000000000005287; Published online 21 September 2023.)*

INTRODUCTION

A recent investigation by Fabi et al[1] has revealed that forehead lines are, among lateral canthal lines and under-eye bags or dark circles, the most concerning facial characteristics when assessing one's own face.[2] Forehead lines are created by contraction of the frontalis muscle, which spans from the eyebrows up to the fronto-occipital transition over the entire forehead. Upon contraction of the frontalis muscle, the cranial part of the forehead moves in a caudally oriented direction, whereas the caudal part of the forehead moves in a cranially oriented direction.[3] Due to this bidirectional movement of the forehead, the cranial and caudal parts of the superficial fatty layer and the skin, which are connected by retinaculae cutis and the suprafrontalis fascia with the frontalis muscle, are pushed against each other and cause a rippling of the forehead.[4] This rippling of the forehead is reflected in the formation of dynamic forehead lines. With increasing age, dynamic forehead lines often become static, which means that even at rest forehead lines can be detected on the forehead. Dynamic forehead lines can be ameliorated using neuromodulators, commonly known as botulinum toxin, or by permanent denervation of the temporal branches of the facial nerve.[5–8] Static forehead lines can be indirectly treated using neuromodulators or denervation, as the decreased movement of the forehead relaxes the skin and causes a progressive amelioration of static forehead lines.[2,6,9] However, in some instances, the static forehead lines are too deep, which requires the use of soft-tissue fillers, most commonly hyaluronic acid–based soft tissue fillers. The presence of forehead lines, static or dynamic, is considered a concerning sign of aging by many individuals. The use of validated photonumeric scales aids in the decision-making of the severity grade of forehead lines.

*From the *Clique Clinic, Kuala Lumpur, Malaysia; †Division of Cosmetic Science, Department of Chemistry, University of Hamburg, Hamburg, Germany; ‡SkinConcept, Munich, Germany; §Private Practice for Dermatology and Aesthetics, Munich, Germany; ¶Department of Plastic Surgery, Lenox Hill Hospital, New York, N.Y.; ||Ocean Clinic, Marbella, Spain; **Yuvell, Aesthetic Clinic & Clinical Trial Center, Vienna, Austria; ††Clinical Research and Development, ICA Aesthetic Navigation GmbH, Frankfurt, Germany; and ‡‡CH Dermatologia, Medellin, Colombia.*

Received for publication June 9, 2023; accepted July 27, 2023.

DOI: 10.1097/GOX.0000000000005287

Disclosure statements are at the end of this article, following the correspondence information.

To the knowledge of the authors, three validated assessment scales for the severity of forehead lines have been published so far. However, the last scale was published in 2016, and most of the available scales have not been validated in the digital and real-life setting.[10–12] Thus, the aim of this investigation was to validate two photonumeric scales for the assessment of static and dynamic forehead lines.

Takeaways

Question: The objective of this investigation was to create and validate five-point photonumeric scales that assess static and dynamic forehead lines.

Findings: The Croma Static Forehead Lines–Assessment Scale and the Croma Dynamic Forehead Lines–Assessment Scale have excellent inter and intra-rater agreements.

Meaning: The Croma Static Forehead Lines–Assessment Scale and the Croma Dynamic Forehead Lines–Assessment Scale are apt to be used in the clinical and study setting, both digitally and live across ethnic groups.

MATERIALS AND METHODS

The creation and validation of the Croma Static Forehead Lines–Assessment Scale and the Croma Dynamic Forehead Lines–Assessment Scale were performed according to a previously published protocol.[13–15]

Creation of a Photobase

A photographer created a database consisting of photographs, which was reviewed on a regular basis by the medical team. Photographs of eligible subjects were taken in a professional photography studio by a trained photographer. A standardized protocol with standardized photographic techniques with a professional, high-resolution color digital camera, and photographic equipment was used. The area to be photographed for assessment of the static and dynamic forehead lines was defined to range from the hairline to the lower eyelid, capturing the entire upper face and parts of the medial face (Fig. 1). Standardized lighting of the photographs was upon discretion of the photographer. Subjects who volunteered to be photographed were selected according to predefined inclusion and exclusion criteria (Table 1).

Fig. 1. Photographed area of interest for the assessment of forehead lines.

Creation of the Grading Scales

A suitable group of photographed subjects was screened from the database by an experienced medical team from H&P Ambulatorien Betriebs GmbH, consisting of dermatologists and plastic surgeons. In the screening process, 122 subject photographs representing every grade of static and dynamic forehead lines were selected to represent diversity in sex and Fitzpatrick skin type per grade (for both, static and dynamic forehead lines). Four independent key opinion leaders in the field of aesthetic

Table 1. Inclusion and Exclusion Criteria for the Respective Scales

Croma Static and Dynamic Forehead Lines–Assessment Scale
Inclusion criteria
Male or female sex, 18 years of age or older
Healthy skin in the face which is free of diseases that could interfere with evaluation
Willingness to abstain from any aesthetic or surgical procedures in the face between the time of photographs and the live evaluation
Written consent that participants might be available for live evaluation after taking the photographs
Written, signed, and dated informed consent
Capable of understanding information about the investigation, including subjects' obligations, and willingness to take part, as evidenced by signed and dated informed consent.
Exclusion criteria
Presence of infectious, inflammatory, or proliferative lesions in the face
Cutaneous lesions in the face
History of major reconstructive surgery (ie, split-thickness grafts or free flaps in the face)
History of major maxillofacial surgery in the face
Tattoos in the face
Permanent make-up in the face
Jewelry of any kind that cannot be removed in the face
Epilepsy
Subjects whose participation in clinical trials is prohibited by the Austrian Medical Devices Act (eg, persons with a legal custodian appointed due to mental disability, prisoners, soldiers and other members of the armed forces, civil servants)

medicine rated the subjects during an online rating. A five-point ordinal rating scale was given to them in addition to an image of the respective condition with absence of static and dynamic forehead lines (grade 0) and an image of very severe static and dynamic forehead lines (grade 5). Based on the mean rating obtained, subjects were then ranked and attributed with their respective rated grade. After the independent rating of the forehead lines photographs, a morphed scale was created, by selecting one subject whose image is representative of general appearance of the condition (grade 3). To create the rest of the gradings for the morphed scale, additional images were selected from the photograph database, and varying degrees of severity of the respective condition were overlaid onto the base. After morphing the scale, a photoguide was created by choosing from the remaining photograph database. A total of four examples per grade was elected for reference. The medical team reviewed the photoguide on several online meetings. The final scale contains the scale descriptors for each grade, an assessment guide, the morphed image, and the real subject images drawn from the digital photoguide (Figs. 2 and 3).

Digital Validation of the Scales

Validation of the scales was conducted in two stages: digitally and live, whereas each stage included two separate rounds of validation. before initiation of the digital validation, all raters, who were key opinion leaders in the aesthetic medical field (see disclosures of this article for further information) participated in an interactive online group training session to familiarize them with the scales, descriptors, photoguide, and rating process. The digital validation process involved the randomization of the subjects' photographs, which were uploaded to the validation online platform (https://studies.yuvell.at/). An approximately equal distribution of severity gradings, based on the initial online-rating, was chosen. The system was a web-based online survey tool in which subjects were registered and managed in a pseudonymous way. Metadata like gender, Fitzpatrick skin type and picture ID, as well as subject photographs to be evaluated were assigned to the single subjects. For the Croma Static Forehead Lines–Assessment Scale, raters were asked to rate the severity of the static forehead lines, and for the Croma Dynamic Forehead Lines–Assessment Scale, raters were asked to grade the severity of the dynamic forehead lines. After a 2-week interval, the digital evaluation was repeated with the same photographs, but in a differently randomized sequence. After each evaluation session, the data were exported in a password-protected excel sheet and sent to the statistician. Before performing the statistical analysis, the statistician verified the correctness of data entries. Statistical analysis was performed after each digital evaluation round was completed and as far as data were available.

Live Validation of the Scale

Approximately 3 months after their photographs were taken, subjects whose images have been used in

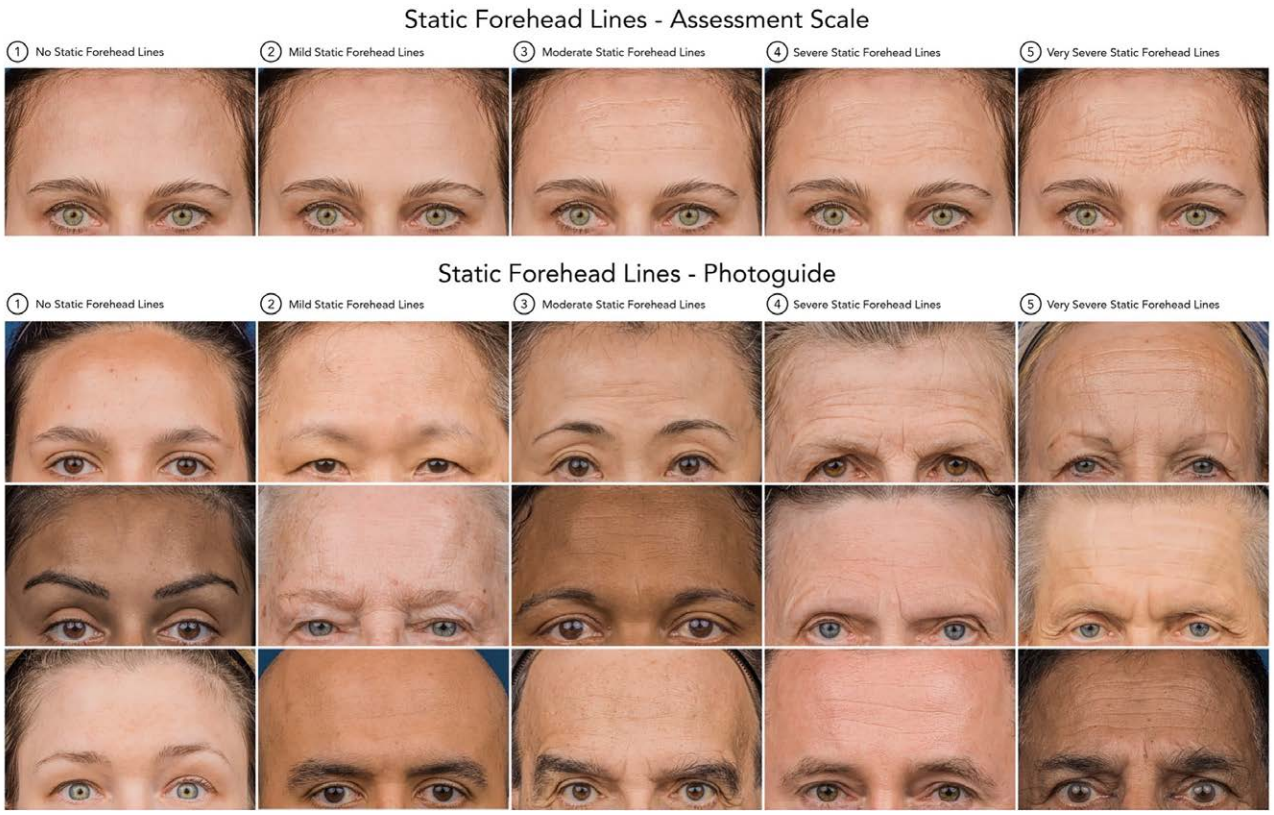

Fig. 2. The Croma Static Forehead Lines–Assessment Scale.

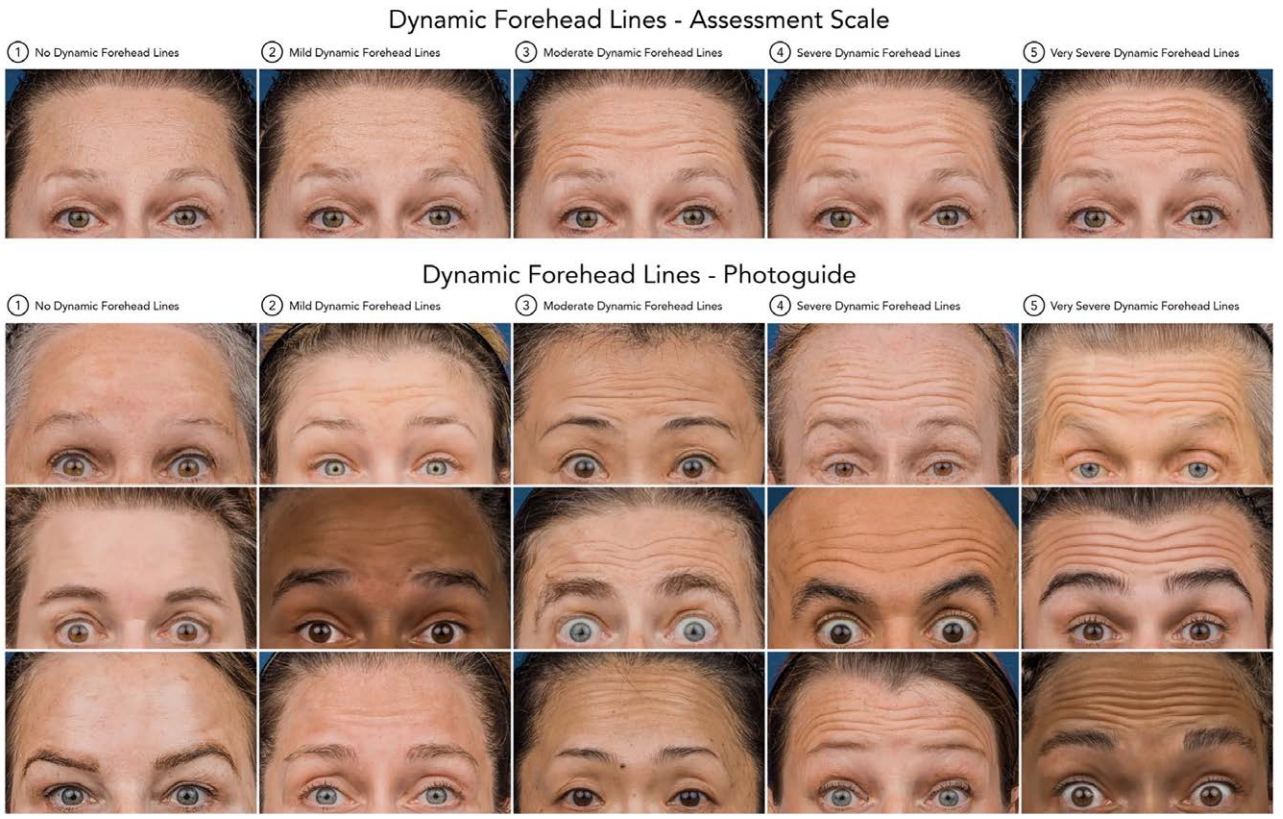

Fig. 3. The Croma Dynamic Forehead Lines–Assessment Scale.

the digital evaluation, but not in the process of scale creation, were invited to participate in two live validation sessions spread over 2 days, with an interval of 2–3 weeks in between.

Subjects were instructed to arrive at the validation site clean, to remove make-up and jewelry, not to drink alcohol excessively before the sessions, to try not to alter their usual routine (eg, their face care routine and normal sleep or hydration patterns) between sessions, and to avoid tanning or extensive sun exposure. Upon arrival at the studio for the first live validation session, subjects signed an informed consent form, and their age, gender, and ethnicity were recorded. Raters were provided with the printed scale in an A3 format. During the first and second live validation sessions, each physician rater evaluated all subjects. Before the first live validation session, all expert raters attended an introductory meeting with the medical lead. Raters had a separate evaluation compartment with an examination lamp (standardized light set up) and the photonumeric scale mounted and displayed for use in subject evaluation. Subjects presented themselves to each rater individually and proceeded from one rating compartment to the next in the same order until being evaluated by all four raters. Raters were instructed to avoid any discussion or conversation with subjects or other raters and switch their rooms every two hours, combined with regular coffee breaks. The raters took at least a 10-minute break every hour and at least a 30-minute lunch break to avoid fatigue.

Sample Size Calculation

Sample size considerations are based on the expected width of the confidence intervals (CIs) for intraclass correlation coefficients (ICCs), with measures of inter-rater reliability being the primary measure of interest. After Bonett, the number of subjects needed to achieve a certain width of the CI depends critically on the number of raters and expected ICC values.[16] Different numbers of raters are planned for live subject validation with five raters and digital subject validation with 10 raters. A range of expected ICC values is derived from guidelines for inter-rater reliabilities in Fleiss, where a value of 0.7 is categorized as "good."[17] Aiming at achieving a (true) reliability of 0.7, confidence widths are derived for inter-rater reliabilities between 0.6 and 0.8 for five and 10 raters, respectively. In case of digital validation with 10 different judges, a sample size of 60 subjects leads to 95% CIs of widths 0.20, 0.17, and 0.13 for overall ICC values of 0.6, 0.7 and 0.8, respectively. Live subject validation is carried out by five judges, and 70 subjects lead to 95% CIs of widths 0.205, 0.17, and 0.13 for overall ICC values of 0.6, 0.7 and 0.8, respectively. Allowing a drop-out rate of 10% after screening, up to 66 and 77 subjects for digital and live validation, respectively, were to be enrolled and screened until the above sample sizes of 60 and 70 subjects were selected and had confirmed their availability to attend the scale evaluations.

Statistical Analysis

The distribution of grades for all raters and photographs (digital validation) and all raters and subjects (live

validation) were provided by frequency counts for round 1 and round 2. Intra-rater reliability between round 1 and round 2 were evaluated for each rater by means of percentage of agreement (exact and ≥1-grade difference) and weighted Cohen kappa statistics using quadratic weights (Fleiss-Cohen weights). ICCs (two-way mixed model without interaction for single measurement) and 95% CIs were derived for weighted kappas and ICCs. Intra-rater reliability was summarized over all raters by calculating means and ranges. Inter-rater reliability was assessed for each round separately by percentage of agreement (exact and ≥1-grade difference) and weighted Cohen's kappa statistics using quadratic weights (Fleiss-Cohen weights) for pairs of raters and means of kappas over all pairs of raters. ICCs (ICC, two-way random model for agreement for single measurement) for pairs of raters and all raters 95% CIs were derived for pairwise and overall weighted kappas and ICCs. Statistical analysis was carried out using the statistical programming language R (version 4.03) including relevant R packages in particular psych.

RESULTS

Croma Static Forehead Lines–Assessment Scale

For the digital rating, a total of 122 subjects, of which 97 (79.5%) were women, with a mean age of 51 wk ± 14.9 years [18–74 years] were assessed. For the live rating, a total of 101 subjects, of which 82 (81.2%) were women, with a mean age of 51.4 ± 15.1 years [18–74 years] were assessed. Further demographics are defined in Table 2. Grading distributions of the digital and live ratings are given in Tables 3 and 4.

Table 2. Demographics of the Subjects Rated in the Digital and Live Validation Rounds for the Created Scales

	Croma Static and Dynamic Forehead Lines–Assessment Scale	
	Digital Validation	Live Validation
No. subjects	122	101
Mean age (SD)	51.2 ± 14.9	51.4 ± 15.1
Minimum	18	18
Maximum	74	74
Women	97 (79.5%)	82 (81.2%)
Fitzpatrick skin type		
I	3 (2.5%)	2 (2.0%)
II	50 (41.0%)	39 (38.6%)
III	43 (35.2%)	38 (37.6%)
IV	14 11.5%)	14 (13.9%)
V	12 (9.8%)	8 (7.9%)

In the digital rating, the inter-rater ICCs were 0.86 (95% CI: 0.82–0.89) in the first rating and 0.82 (95% CI: 0.78–0.86) in the second rating, while the kappa value was 0.86 (95% CI: 0.83–0.988) for the first rating and 0.82 (95% CI: 0.78–0.86) for the second rating. The intra-rater ICCs were 0.88 (95% CI: 0.84–0.91) with a kappa value of 0.87 (95% CI: 0.84–0.91) (Table 5).

In the live rating, the inter-rater ICCs were 0.81 (95% CI: 0.71–0.88) in the first rating and 0.82 (95% CI: 0.75–0.87) in the second rating and a kappa value of 0.81 (95% CI: 0.77–0.85) in the first rating and 0.82 (95% CI: 0.77–0.86) in the second rating. The intra-rater ICCs were 0.87 (95% CI: 0.81–0.93) with a kappa value of 0.87 (95% CI: 0.80–0.93) (Table 6).

Croma Dynamic Forehead Lines–Assessment Scale

For the digital rating, a total of 122 subjects, of which 97 (79.5%) were women, with a mean age of 51.1 ± 14.8 years (18–74 years) were assessed. For the live rating, a total of 101 subjects, of which 82 (81.2%) were women, with a mean age of 51.4 ± 15.1 years (18–74 years) were assessed. Further demographics are defined in Table 2. Grading distributions of the digital and live ratings are given in Table 3 and 4.

In the digital rating, the inter-rater ICCs were 0.83 (95% CI: 0.79–0.86) in the first rating and 0.80 (95% CI: 0.75–0.84) in the second rating, while the kappa value was 0.83 (95% CI: 0.79–0.87) for the first rating and 0.80 (95% CI: 0.75–0.83) for the second rating. The intra-rater ICCs were 0.85 (95% CI: 0.80–0.90) with a kappa value 0.85 (95% CI: 0.80–0.89) (Table 5).

In the live rating, the inter-rater ICCs were 0.71 (95% CI: 0.63–0.79) in the first rating and 0.78 (95% CI: 0.72–0.84) in the second rating and a kappa value of 0.71 (95% CI: 0.64–0.77) in the first rating and 0.78 (95% CI: 0.72–0.83) in the second rating. The intra-rater ICCs were 0.80 (95% CI: 0.73–0.86) with a kappa value of 0.80 (95% CI: 0.73–0.86) (Table 6).

DISCUSSION

The created scales have been developed for the assessment of static and dynamic forehead lines in both sexes, all age groups, and different Fitzpatrick skin types. The Croma Static Forehead Lines–Assessment Scale has shown almost perfect inter-rater and intra-rater agreement in both the digital and the live validations, while the Croma Dynamic Forehead Lines–Assessment Scale has shown almost perfect inter and intra-rater agreement in the digital validation. In the live validation,

Table 3. Number of Rated Gradings (Including %) in the Digital Rating for the Respective Created Scales

	Croma Static Forehead Lines–Assessment Scale		Croma Dynamic Forehead Lines–Assessment Scale	
	First Session	Second Session	First Session	Second Session
Grade 1 (none)	230 (17.1%)	253 (18.9%)	72 (5.4%)	77 (5.7%)
Grade 2 (minimal)	506 (37.7%)	476 (35.5%)	405 (30.2%)	340 (25.3%)
Grade 3 (moderate)	281 (20.9%)	298 (22.2%)	371 (27.6%)	350 (26.1%)
Grade 4 (severe)	207 (15.4%)	212 (15.8%)	365 (27.2%)	387 (28.8%)
Grade 5 (very severe)	118 (8.8%)	103 (7.7%)	129 (9.6%)	188 (14.0%)

Table 4. Number of Rated Gradings (Including %) in the Live Rating for the Respective Created Scales

	Croma Static Forehead Lines–Assessment Scale		Croma Dynamic Forehead Lines–Assessment Scale	
	First Session	Second Session	First Session	Second Session
Grade 1 (none)	58 (14.4%)	51 (12.6%)	11 (2.7%)	9 (2.2%)
Grade 2 (minimal)	151 (37.5%)	157 (38.9%)	114 (28.3%)	94 (23.3%)
Grade 3 (moderate)	99 (24.6%)	83 (20.5%)	131 (32.5%)	128 (31.7%)
Grade 4 (severe)	63 (15.6%)	68 (16.8%)	99 (24.6%)	122 (30.2%)
Grade 5 (very severe)	32 (7.9%)	45 (11.1%)	48 (11.9%)	51 (12.6%)

Table 5. ICCs and Kappas for the Digital Rating and Live Rating of the Croma Static Forehead Lines–Assessment Scale and Their Respective Interpretation according to Fleiss et al, as well as the ICCs and Kappas of Comparable Scales

	Croma Static Forehead Lines–Assessment Scale	Interpretation	Carruthers et al	Interpretation	Flynn et al	Interpretation	Carruthers et al	Interpretation
Digital rating								
Inter-rater ICC	0.86 [95% CI: 0.82–0.89] and 0.82 [95% CI: 0.78–0.86]	Almost perfect	0.846 and 0.863	Almost perfect	0.67 and 0.65	Substantial	—	—
Inter-rater kappa	0.86 [95% CI: 0.83–0.88] and 0.82 [95% CI: 0.78– 0.86]	Almost perfect	—	—	—	—	—	—
Inter-rater ICC	0.88 [95% CI: 0.84–0.91]	Almost perfect	0.846–0.942	Almost perfect	0.82 (0.72–0.91)	Substantial	—	—
Inter-rater kappa	0.87 [95% CI: 0.84–0.91]	Almost perfect			—	—	—	—
Live rating								
Inter-rater ICC	0.81 [95% CI: 0.71–0.88] and 0.82 [95% CI: 0.75–0.87]	Almost perfect	—	—	—	—	0.87 (0.818–0.916) and 0.86 (0.815–0.908)	Almost perfect
Inter-rater kappa	0.81 [95% CI: 0.77–0.85] and 0.82 [95% CI: 0.77–0.86]	Almost perfect	—	—	—	—		—
Intra-rater ICC	0.87 [95% CI: 0.81–0.93]	Almost perfect	—	—	—	—	0.87 (0.781–0.952)	Almost perfect
Intra-rater kappa	0.87 [95% CI: 0.80–0.93]	Almost perfect	—	—	—	—		

Table 6. ICCs and Kappas for the Digital Rating and Live Rating of the Croma Dynamic Forehead Lines–Assessment Scale and Their Respective Interpretation According to Fleiss et al, as well as the ICCs and Kappas of Comparable Scales

	Croma Dynamic Forehead Lines–Assessment Scale	Interpretation	Carruthers et al	Interpretation	Flynn et al	Interpretation
Digital rating						
Inter-rater ICC	0.83 [95% CI: 0.79–0.86] and 0.80 [95% CI: 0.75–0.84]	Almost perfect	0.852 and 0.892	Almost perfect	0.70 and 0.63	Substantial
Inter-rater kappa	0.83 [95% CI: 0.79–0.87] and 0.80 [95% CI: 0.75– 0.83]	Almost perfect	—	—	—	—
Intra-rater ICC	0.85 [95% CI: 0.80–0.90]	Almost perfect	0.859–0.941	Almost perfect	0.80 (0.72–0.90)	Almost perfect
Intra-rater kappa	0.85 [95% CI: 0.80–0.89]	Almost perfect	—	—	—	—
Live rating						
Inter-rater ICC	0.71 [95% CI: 0.63–0.79] and 0.78 [95% CI: 0.72–0.84]	Substantial		—	—	—
Inter-rater kappa	0.71 [95% CI: 0.64–0.77] and 0.78 [95% CI: 0.72–0.83]	Substantial	—	—	—	—
Intra-rater ICC	0.80 [95% CI: 0.73–0.86]	Almost perfect	—	—	—	—
Intra-rater kappa	0.80 [95% CI: 0.73–0.86]	Almost perfect	—	—	—	—

substantial inter-rater agreement and almost perfect intra-rater agreement has been shown. Assessment of forehead lines can be considered very difficult, as the number of forehead lines is highly variable, and at the same time there is a plethora of contraction patterns, which might obscure a precise assessment. The created scales for both static and dynamic forehead lines have been shown to be a reliable guidance in the assessment, based on their inter and intra-rater agreement. The obtained inter- and intra-rater agreements are in line with the values of previously published scales.[10–12] The created scale for static forehead lines revealed higher intra and inter-rater agreement values than any other reported scale. At the same time, the created scale for dynamic forehead lines achieved at least comparable or better inter- and intra-rater agreement. Interestingly, the inter- and intra-rater agreement for assessment of static forehead lines has been greater than for the assessment of dynamic forehead lines, even though upon contraction of the frontalis muscle the number of forehead lines created and their contraction pattern varies greatly. A unique property of the created scales is their validation in the digital and live setting. Of all previously published scales, none has been validated in both settings, but rather, either digitally only or live only. With increasing regulatory requirements, the efficacy of medical devices, as well as drugs often needs to be assessed in the live setting by a treating or blinded investigator and subject, but also by an independent review panel, which needs to assess the improvement in a digital manner (ie, by rating the severity of forehead lines on a digital photograph). Thus, it is important to have scales available, which have been validated in the digital and in the live setting. Although scales classically originate from the clinical study setting, have been the only tool to measure efficacy in aesthetic medicine for a long time, and are still considered the gold standard, the use of scales has also gained popularity in the clinical setting.[13–15] By showing a patient his initial forehead line severity upon rest (static) and eye-brow elevation (dynamic) a physician can properly discuss the possibilities and the anticipated outcome of a treatment with the patient, which facilitates communication overall. At the same time, it is a useful tool to manage expectations with the patient, as very severe static forehead lines will not become mild static forehead lines with the use of botulinum toxin within a time frame of a couple of weeks. Moreover, it allows proper documentation of the starting point before an intervention with neuromodulators or soft-tissue fillers. At the same time, photonumeric scales pose a certain limitation to the objective assessment of a condition, which is reflected by inter- and intra-rater agreements smaller than 1.0. The reason for this is the human being that performs the assessment. In the foreseeable future, the assessment of aesthetic conditions will most likely be at least supported by artificial intelligence (AI), which can use adjunct tools as three-dimensional surface imaging or surface electromyography for a more precise assessment. Combining facial aesthetic scales with AI has the potential to offer several benefits in the future. Facial aesthetic scales provide a structured framework for evaluating facial features and assessing aesthetic changes. By incorporating AI, which can analyze facial images and measurements, it becomes possible to automate the process of rating and quantifying aesthetic attributes. This helps establish standardized assessments that are consistent and objective across different practitioners, research studies, or clinical settings.

Furthermore, AI algorithms can analyze facial features and textures with high precision and consistency. By leveraging machine learning techniques, AI systems can be trained on large datasets of facial images, aesthetic ratings, and expert evaluations. This enables them to learn patterns and correlations between facial attributes and aesthetic scales, leading to improved accuracy and reliability in assessing facial aesthetics. In terms of personalized treatment planning, AI systems can analyze large datasets of facial aesthetics and treatment outcomes to identify patterns and correlations. By combining facial aesthetic scales with AI, it becomes possible to develop predictive models that can help in personalized treatment planning. For example, AI algorithms can analyze a patient's facial features, compare them to a reference database, and provide recommendations on the most suitable cosmetic procedures or treatments to achieve desired aesthetic outcomes.

Apart from guidance in the clinical treatment, facial aesthetic scales, when combined with AI, can facilitate longitudinal tracking and monitoring of facial changes over time. By regularly capturing facial images and using AI algorithms to assess aesthetic features, practitioners can objectively measure the effectiveness of cosmetic interventions, track progression or regression of facial aging, and make data-driven decisions regarding treatment plans.

It is important to note that although AI can enhance the assessment process, it should not replace the expertise and judgment of healthcare professionals. AI systems should be considered as tools to assist and augment human decision-making, rather than as standalone evaluators. Close collaboration between AI algorithms and human practitioners is crucial to ensure the accurate and ethical use of facial aesthetic scales in combination with AI. It remains elusive when assessment by human beings of the severity of an aesthetic condition will be replaced by AI; however, for now, the gold standard remains the use of photonumeric scales.

CONCLUSIONS

Facial aesthetic scales are beneficial in clinical practice because they provide a standardized framework for assessing and documenting aesthetic changes, aiding in communication with patients, planning treatment, and monitoring treatment progress. Additionally, they support evidence-based practice and research by facilitating data collection and comparison of treatment outcomes. The Croma Static Forehead Lines–Assessment Scale and the Croma Dynamic Forehead Lines–Assessment Scale have

excellent inter- and intra-rater agreements to be justifiably used in the clinical and study setting, both digitally and live across ethnic groups.

Tingsong Lim, MD
Clique Clinic
BO3A-06-01, Menara 3A, 3 Jalan Bangsar
KL Eco City, Kuala Lumpur 59200
Malaysia
E-mail: drlim@cliqueclinic.com

DISCLOSURES

The authors have no financial interest to declare in relation to the content of this article. This study received funding from Croma Pharma GmbH, Leobendorf, Austria. The authors were financially compensated for the efforts taken to validate the scales.

REFERENCES

1. Fabi S, Alexiades M, Chatrath V, et al. Facial aesthetic priorities and concerns: a physician and patient perception global survey. *Aesthet Surg J.* 2022;42:NP218–NP229.
2. Frank K, Freytag DL, Schenck TL, et al. Relationship between forehead motion and the shape of forehead lines—a 3D skin displacement vector analysis. *J Cosmet Dermatol.* 2019;18:1224–1229.
3. Cotofana S, Freytag DL, Frank K, et al. The bidirectional movement of the frontalis muscle: introducing the line of convergence and its potential clinical relevance. *Plast Reconstr Surg.* 2020;145:1155–1162.
4. Davidovic K, Melnikov D, Frank K, et al. To click or not to click—The importance of understanding the layers of the forehead when injecting neuromodulators—A clinical, prospective, interventional, split-face study. *J Cosmet Dermatol.* 2021;20:1385–1392.
5. de Maio M, Swift A, Signorini M, et al; Aesthetic Leaders in Facial Aesthetics Consensus Committee. Facial assessment and injection guide for botulinum toxin and injectable hyaluronic acid fillers: focus on the upper face. *Plast Reconstr Surg.* 2017;140:265e–276e.
6. Iozzo I, Tengattini V, Antonucci VA. Multipoint and multilevel injection technique of botulinum toxin A in facial aesthetics. *J Cosmet Dermatol.* 2014;13:135–142.
7. Small R. Botulinum toxin injection for facial wrinkles. *Am Fam Physician.* 2014;90:168–175.
8. Abramo AC, Sgarbi R, Scartozzoni M. Histological and clinical appraisal of distribution and long acting new hyaluronic acid into the subcutaneous fat of the forehead to improve horizontal lines. *Aesthetic Plast Surg.* 2021;45:289–295.
9. Carruthers A, Donofrio L, Hardas B, et al. Development and validation of a photonumeric scale for evaluation of static horizontal forehead lines. *Dermatol Surg.* 2016;42(Suppl 1):S243–S250.
10. Carruthers A, Carruthers J, Hardas B, et al. A validated grading scale for forehead lines. *Dermatol Surg.* 2008;34(Suppl 2):S179–S183.
11. Flynn TC, Carruthers A, Carruthers J, et al. Validated assessment scales for the upper face. *Dermatol Surg.* 2012;38:309–319.
12. Pavicic T, Pooth R, Prinz V, et al. Validated 5-point photonumeric scales for the assessment of the periorbital region. *J Cosmet Dermatol.* 2022;21:158–166.
13. Pooth R, Prinz V, Cajkovsky M, et al. Validated 5-point photonumeric scales for the assessment of the jowls and chin. *J Cosmet Dermatol.* 2022;21:600–607.
14. Green JB, Pavicic T, Pooth R, et al. Validated 5-point photonumeric scales for the assessment of hand atrophy. *J Cosmet Dermatol.* 2022;21:933–939.
15. Bonett DG. Sample size requirements for testing and estimating coefficient alpha. *J Educ Behav Stat.* 2002;27:335–340.
16. Fleiss JL, Levin B, Paik MC. The measurement of interrater agreement. In: *Statistical Methods for Rates and Proportions.* 2004.
17. Cotofana S, Sattler S, Frank K, et al. Aesthetic medicine—quo vadis? *J Cosmet Dermatol.* 2022;21:6526–6527.