



# A novel photonumeric hand grading scale for hand rejuvenation

Jong Hun Lee<sup>1</sup>, Yean Su Choi<sup>2</sup>, Eun Soo Park<sup>2</sup>, Jong Seo Kim<sup>3</sup>, Moon Seok Kang<sup>4</sup>, Hwa Young Oh<sup>5</sup>, So Dam Yang<sup>6</sup>, Seon Hui Jeon<sup>6</sup>

<sup>1</sup>Department of Plastic and Reconstructive Surgery, Nowon Eulji Medical Center, Eulji University School of College, Seoul; <sup>2</sup>Department of Plastic and Reconstructive Surgery, Soonchunhyang University Bucheon Hospital, Bucheon; <sup>3</sup>Kim Jongseo Plastic Surgery Clinic, Seoul; <sup>4</sup>Garosu Plastic and Aesthetic Surgery Clinic, Seoul; <sup>5</sup>Jayjun Plastic and Aesthetic Surgery Clinic, Seoul; <sup>6</sup>Life Sciences R&D, LG Chem, Ltd., Seoul, Korea

Original Article

**Background** Few scales are currently available to evaluate changes in hand volume. We aimed to develop a hand grading scale for quantitative assessments of dorsal hand volume with additional consideration of changes in skin texture; to validate and prove the precision and reproducibility of the new scale; and to demonstrate the presence of clinically significant differences between grades on the scale.

**Methods** Five experienced plastic surgeons developed the Hand Volume Rating Scale (HVRS) and rated 91 images. Another five plastic surgeons validated the scale using 50 randomly selected images. Intra- and inter-rater agreement was calculated using the weighted kappa statistic and intraclass correlation coefficients (ICCs). Paired images were also evaluated to verify whether the scale reflected clinical differences.

**Results** The intra-rater agreement was 0.95 (95% confidence interval, 0.922–0.974). The inter-rater ICCs were excellent (first rating, 0.94; second rating, 0.94). Image pairs that differed by 1, 2, and 3 grades were considered to contain clinically relevant differences in 80%, 100%, and 100% of cases, respectively, while 84% of image pairs of the same grade were found not to show clinically relevant differences. This confirmed that the scale of the HVRS corresponded to clinically relevant distinctions.

**Conclusions** The scale was proven to be precise, reproducible, and reflective of clinical differences.

**Keywords** Asian continental ancestry group / Back / Hand / Rejuvenation / Skin aging

**Correspondence:** Eun Soo Park  
Department of Plastic and Reconstructive Surgery,  
Soonchunhyang University Bucheon  
Hospital, 170 Jomaru-ro, Wonmi-gu,  
Bucheon 14584, Korea  
Tel: +82-32-621-5319  
Fax: +82-32-621-5662  
E-mail: peunsoo@schmc.ac.kr

This work was supported by Life Sciences R&D, LG Chem, Ltd.

The authors acknowledge and thank the following members of the scale validation group for completing the scale validation: Kook Hyun Kim, MD, Byungkwan Shim, MD, Han Jeong Lee, MD, Seok Min Choi, MD, and Han Gyu Cha, MD.

Received: 6 May 2019 • Revised: 10 Jul 2019 • Accepted: 10 Jul 2019  
pISSN: 2234-6163 • eISSN: 2234-6171 • <https://doi.org/10.5999/aps.2019.00521> • Arch Plast Surg 2019;46:359-364

## INTRODUCTION

Recently, increased interest has emerged in hand rejuvenation to obtain a more youthful appearance. Treatments using hyaluronic acid filler, calcium hydroxyapatite filler, poly-L-lactic acid

filler, or botulinum toxin have become widely used for the recovery of hand volume and hand rejuvenation, and scales for objectively evaluating the improvements yielded by such techniques are being developed.

Carruthers et al. [1,2] and Jones et al. [3] developed grading

scales for evaluating hand volume changes; however, those scales were primarily developed based on the hands of individuals with Fitzpatrick skin type II, implying that they are mainly applicable to Caucasians, and focused on defining hand volume changes. To evaluate the recovery of hand volume and the outcomes of hand rejuvenation procedures, a new rating scale is needed that would evaluate the quantity of hand volume and specifically consider changes in skin texture due to hand volume changes. Such a scale could be used for clinical guidelines, and would enable a standard and objective evaluation of clinical trial outcomes.

First, we planned to develop a 5-grade photonumeric scale assessing the volume of the dorsal hand and to establish a validation process to demonstrate the precision and reproducibility of the new scale. Second, we aimed to demonstrate the clinical significance of differences between grades on the scale. Finally, according to the protocol, a new hand grading scale composed of photographs representing each grade and text descriptions explaining each grade was developed and validated.

## METHODS

### Study design

A total of 164 subjects met the inclusion criteria (age,  $\geq 20$  years; race, Asian [Korean]). In addition, the subjects' right hands did not have any severe scars, wounds, tattoos, or excessive hair that could potentially influence the score on the grading scale. All photographs were obtained under the same settings (camera type, Canon EOS 700D, Tokyo, Japan; lens type, EF 50 mm f/1.8; distance from lens to object, 91 cm; shutter speed, 1/125; aperture, F14; ISO, 100; light: Poton-150).

The protocol of this study was approved by the ethics committee of Soonchunhyang University Bucheon Hospital (IRB No. 2019-05-012-002). Subjects were provided with an explanation of this study, and they gave written consent for use of their photographs.

### Scale development

The new hand grading scale, the Hand Volume Rating Scale (HVRS), is a 5-grade photonumeric rating scale for objectively evaluating hand volume changes in Asians. Five experienced plastic surgeons (JHL, ESP, JSK, MSK, and HYO) were included in the scale development group. The dorsal portion of the hand was defined as the area from the metacarpophalangeal joints to 1 cm distal to the wrist based on the definition of Canfield [3]. The text descriptions of each grade are as follows: 0 (absent), no soft tissue loss, no visible or only superficially visible veins, and no visible tendons; 1 (minimal), minimal soft tis-

sue loss, slightly prominent veins, and no or barely visible tendons; 2 (moderate), moderate soft tissue loss, prominent veins, and markedly visible tendons; 3 (moderately severe), moderately severe soft tissue loss, very prominent veins, substantially protruding tendons (most tendons are visible), and rough skin (the presence of fine wrinkles) (all of the aforementioned conditions are required for grade 3); and 4 (severe), severe soft tissue loss, pronouncedly prominent veins, extremely protruding tendons (all tendons are visible), and severely rough skin with severe dermal atrophy (severe presence of fine wrinkles) (all of the aforementioned conditions are required for grade 4). Visible veins are usually located in the intermediate fatty lamina and the dorsal intermediate lamina. Deep to this layer, the dorsal deep lamina contains the extensor tendons. Finally, the dorsal interosseous muscles and metacarpal bones are covered by the dorsal deep fascia [4]. Loss of subcutaneous tissue within each lamina and muscle atrophy cause several of the features of hand volume loss, such as protruding veins, protruding tendons, and skin redundancy. Thus, the descriptions of the lower grades include superficial changes in terms of veins, tendons, and skin texture.

Based on the pre-defined image exclusion criteria, such as incorrect posture, incorrect photograph settings, and hands with scarring or skin disease, 91 images in total were selected and included in the photographic database after excluding inappropriate images from 73 subjects. After assessing the grades of each image, the scale development group selected two representative images for each grade that corresponded closely to the text description. Using one of the representative images of grade 2 as a base image, an external graphic designer drew morphed images to match the descriptions for each grade; these images were reviewed by the scale development group. The text description, the morphed image describing each grade, and two representative real images were produced to facilitate accurate user comprehension (Fig. 1).

### Training

The scale validation group, which was responsible for rating images to validate the scale, included five other plastic surgeons. Each rater in this group underwent a training process before the validation process. For training, both a training booklet including four images per grade, which were selected from the photographic database, and the aforementioned two representative images for each grade were used. Each rater completed face-to-face training with a member of the scale development group and finished individual self-training with the training booklet. To confirm that the training had been successfully completed, each rater had to pass an individual training test before rating the images for validation of the scale.

Fig. 1. The Hand Volume Rating Scale



Confidential

**Scale validation**

After the training booklet was created and representative images were selected, 50 other images were randomly selected for the validation booklet from the photographic database, with an equal distribution of all grades, using SAS version 9.4 (SAS Institute, Cary, NC, USA). Using Bonett’s method [5] for estimating inter-class correlations with the desired precision, it was determined that a sample size of 50 images would be needed to achieve a 95% confidence interval (CI) with a width of 0.2 with five raters. All raters independently performed evaluations of the 50 randomly ordered and blinded images twice. There was at least a 1-week interval between the first and second evaluations.

To determine intra-rater agreement, each rater’s data were assessed by the mean weighted kappa statistic and a corresponding 95% CI calculated using the bootstrapping method. The weighted kappa for each rater was measured using Fleiss-Cohen weighting [6]. The weighted kappa values were interpreted as follows: lower than 0.0, poor agreement; 0.0–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and higher than 0.80, almost perfect agreement [7]. For each evaluation, the inter-class correlation coefficient (ICC) from a two-way mixed effect model, including the rater as a random effect, was calculated according to the Shrout and Fleiss model [8] to assess inter-rater agreement. A 95% CI for each ICC was calculated using the bootstrapping method. The ICC calculated for the second evaluation was considered the primary endpoint of inter-rater agreement, with an ICC lower than 0.40 indicating poor agreement, 0.40–0.60 indicating fair agreement, 0.60–0.75 indicating good agreement, and higher than 0.75 indicating excellent agreement [9]. Therefore, the acceptance criteria for validating this scale as reliable and meaningful were a mean weighted kappa higher than 0.6 and an ICC higher than 0.7 for the second evaluation. For all statistical analyses, SAS version 9.4 (SAS Institute) was used.

To verify whether differences in the grades were clinically significant, paired images were evaluated after assessments of intra-rater and inter-rater agreement. Pairs of photos from the valida-

tion booklet were generated, comprising a total of 32 pairs (10 pairs of the same grade, 12 pairs with a 1-grade difference, 6 pairs with a 2-grade difference, and 4 pairs with a 3-grade difference). Raters answered “yes” or “no” according to their judgment as to whether there was a clinically relevant difference between the paired photos. Then, the proportions of “yes” and “no” responses for each grade difference were calculated. A proportion of “yes” responses exceeding 80% for image pairs with differences of 1, 2, or 3 grades was established as clinically meaningful. In contrast, for the image pairs of the same grade, a proportion of “no” responses exceeding 80% was utilized as the threshold for an absence of clinically meaningful differences between those images.

**RESULTS**

The demographic characteristics of the subjects whose photographs were included in the final scale validation set are shown in Table 1.

The intra-rater agreement assessment showed high consistency within raters, as demonstrated by high weighted kappa scores ranging from 0.91 to 0.98. The mean weighted kappa was 0.95 (95% CI, 0.922–0.974), indicating almost perfect agreement (Table 2). The inter-rater agreement analysis for validation was also meaningful, with an ICC of 0.94 for both the first and second evaluations (Table 3). The aforementioned results satisfied the pre-defined acceptance criteria.

For image pairs with differences of 1, 2, and 3 grades, the proportions of responses indicating that there was a clinically meaningful difference were 80%, 100%, and 100%, respectively,

**Table 1. Demographics of subjects whose photographs were used in the scale validation set**

Variable	Value (n = 91)
Age (yr)	
Mean ± SD	50.4 ± 20.60
Median (range)	41.0 (24–87)
Sex, no. (%)	
Male	17 (18.7)
Female	74 (81.3)
Race, no. (%)	
Asian (Korean)	91 (100.0)

**Table 2. Intra-rater agreement of the Hand Volume Rating Scale**

	κ-value (95% CI)
Rater 1	0.91 (0.856–0.964)
Rater 2	0.96 (0.930–0.988)
Rater 3	0.98 (0.966–1.000)
Rater 4	0.92 (0.874–0.961)
Rater 5	0.97 (0.945–0.995)
Mean weighted κ	0.95 (0.922–0.974)

CI, confidence interval.

**Table 3. Inter-rater agreement of the Hand Volume Rating Scale**

	ICC (95% CI)
Evaluation 1	0.94 (0.900–0.962)
Evaluation 2	0.94 (0.902–0.965)

ICC, intraclass correlation coefficient; CI, confidence interval.

**Table 4. Difference in scores for image pairs by grade difference using the Hand Volume Rating Scale**

Grade difference	Response	Percent (%)
0	No	84.0
1	Yes	80.0
2	Yes	100.0
3	Yes	100.0

confirming that differences  $\geq 1$  on the HVRS were clinically significant. In contrast, 84% of responses indicated that there was no clinically relevant difference between image pairs of the same grade, meaning that the differences between those images were not clinically significant (Table 4).

## DISCUSSION

Few validated scales are presently used for assessing changes in hand volume. Carruthers et al. [1,2] proposed a novel 5-point photonumeric scale for objective quantification of the severity of hand aging. Their scale considered the degree of fatty tissue loss and the visibility of veins and tendons as critical factors in the rating. Morphed images and one untouched validated image of hands for each grade were attached to the grading scale.

Jones et al. [3] reported a new hand grading scale, the Allergan Hand Volume Deficit Scale, for similar purposes. The description of this hand grading scale includes information about tendons and veins. To minimize subjective expressions, the description of the scale selected the words “protruding” or “prominent,” rather than “mild” or “moderate.” The Allergan Hand Volume Deficit Scale showed high inter-rater and intra-rater agreement among physicians. However, neither of those scales included a description of skin texture, which is one of the most remarkable features of the aging process. The HVRS proposed in this study includes the words “rough skin” and “severely rough skin with severe dermal atrophy” for grades 3 and 4, respectively, to clarify the differences between grades.

In this study, a validation process was conducted for the HVRS that was similar those previously used for other scales, demonstrating almost perfect intra-rater and inter-rater agreement for the HVRS. The mean intra-rater weighted kappa of the five raters was 0.95, and the ICC between raters was 0.94. These values are higher than those reported by Carruthers et al. [1,2] and Jones et al. [3]. The high kappa value indicates that the HVRS can be used consistently by the same rater at different times for evaluating hand volume changes, while the high ICC demonstrates that the HVRS can be used consistently by different raters at different times for hand evaluation. In addition, a 1-grade difference was shown to reflect clinically significant differences

in the appearance of the dorsal hands. Therefore, the validation process conducted herein demonstrated that the 5-point photonumeric HVRS can be considered a reliable method for classifying the volume of the dorsal hands in clinical studies.

The detailed text descriptions of the HVRS account for both changes in volume and differences in skin texture due to hand volume changes, which may be one of the most important factors that contributed to the high intra-rater and inter-rater agreement. The scale development group strived to avoid subjective and uniform expressions, and instead selected words that represented specific features of each grade. Both the loss of soft tissue and the status of veins and tendons were described in detail to clarify the clinical differences between grades. A grade of 0 on the HVRS means that the ideal dorsal hand has no visible soft tissue loss, fine wrinkles, veins, or tendons, which is the goal of fillers or botulinum toxin treatment. In addition, the descriptions of grades 3 and 4 included expressions describing skin texture characteristics caused by hand volume changes, such as “fine wrinkles” and “severe dermal atrophy.” These characteristics make the HVRS specific, because it can be used to assess hand volume changes in Asians and accurately reflects changes in hand volume. Therefore, the HVRS should be considered as the most suitable available option for evaluating the effects of hand rejuvenation after treatment with fillers or botulinum toxin.

Although this study showed sufficiently meaningful results, it has some limitations regarding the use of this scale alone for evaluating hand changes due to aging. Our descriptions only deal with soft tissue, veins, tendons, and skin texture. Younger and more attractive hands have supple skin and soft tissue with no wrinkles. In contrast, aging hands have more prevalent wrinkles, thin skin, age spots, prominent veins, more visible tendons, and bony deformities [10]. Age spots are one of the most important factors associated with age, and Jakubietz et al. [11] showed a strong and positive correlation between age and the number of age spots. A description of age spots can be helpful for evaluating aging hands. Furthermore, the hand aging process includes motion effects and skin aging, as well as volume loss and its sequelae [11]. Therefore, a more comprehensive approach that considers other features of aging hands is required for a comprehensive objective evaluation of aging hands. As a special case, many athletic people have numerous veins and well-developed tendons regardless of their age and hand volume. Thus, the hands of such individuals may be difficult to evaluate using the HVRS. In addition, all subjects who participated in this study were Asians; therefore, we could not consider various Fitzpatrick skin types. In the future, the HVRS should be applied to other Fitzpatrick skin types to validate this scale as a more globally accepted parameter.

The almost perfect intra-rater weighted kappa scores of the five raters and the very high ICCs between raters confirmed that the HVRS is a reliable grading scale for assessing hand volume changes in Asians. In addition, because a 1-point score difference reflected clinically significant differences in hand volume, the HVRS is expected to play a useful role in clinical studies.

## NOTES

### Conflict of interest

SDY and SHJ are employees of LG Chem, Ltd. The other authors report no conflict of interest. The LG Hand Grading Scale is owned by Life Sciences R&D, LG Chem, Ltd.

### Ethical approval

The study was approved by the Institutional Review Board of Soonchunhyang University Bucheon Hospital (IRB No. 2019-05-012-002) and performed in accordance with the principles of the Declaration of Helsinki. Written informed consents were obtained.

### Patient consent

The patients provided written informed consent for the publication and the use of their images.

### Author contribution

Conceptualization: Lee JH, Park ES, Kim JS. Data curation: Jeon SH. Formal analysis: Jeon SH. Funding acquisition: Lee JH, Park ES. Methodology: Lee JH, Park ES, Kim JS, Kang MS, Oh HY, Yang SD. Project administration: Yang SD. Visualization: Yang SD. Writing - original draft: Choi YS. Writing - review & editing: Lee JH, Park ES, Choi YS, Yang SD, Jeon SH. Approval of final manuscript: all authors.

### ORCID

Jong Hun Lee <https://orcid.org/0000-0001-6417-6986>

Yean Su Choi <https://orcid.org/0000-0002-7670-1722>

Eun Soo Park <https://orcid.org/0000-0003-2966-9122>

Jong Seo Kim <https://orcid.org/0000-0003-1428-5898>

Moon Seok Kang <https://orcid.org/0000-0001-7072-6091>

Hwa Young Oh <https://orcid.org/0000-0001-5971-9093>

So Dam Yang <https://orcid.org/0000-0003-3376-0236>

Seon Hui Jeon <https://orcid.org/0000-0003-1152-4267>

## REFERENCES

1. Carruthers A, Carruthers J, Hardas B, et al. A validated hand grading scale. *Dermatol Surg* 2008;34 Suppl 2:S179-83.
2. Cohen JL, Carruthers A, Jones DH, et al. A randomized, blinded study to validate the Merz Hand Grading Scale for use in live assessments. *Dermatol Surg* 2015;41 Suppl 1:S384-8.
3. Jones D, Donofrio L, Hardas B, et al. Development and validation of a photonumeric scale for evaluation of volume deficit of the hand. *Dermatol Surg* 2016;42 Suppl1:S195-S202.
4. Bidic SM, Hatf DA, Rohrich RJ. Dorsal hand anatomy relevant to volumetric rejuvenation. *Plast Reconstr Surg* 2010;126:163-8.
5. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21:1331-5.
6. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973;33:613-9.
7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
8. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-8.
9. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284-90.
10. Coleman SR. Hand rejuvenation with structural fat grafting. *Plast Reconstr Surg* 2002;110:1731-44.
11. Jakubietz RG, Kloss DF, Gruenert JG, et al. The ageing hand: a study to evaluate the chronological ageing process of the hand. *J Plast Reconstr Aesthet Surg* 2008;61:681-6.