Clinical, Cosmetic and Investigational Dermatology

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

A software program designed to educate patients on age-related skin changes of facial and exposed extrafacial regions: the results of a validation study

This article was published in the following Dove Press journal:
Clinical, Cosmetic and Investigational Dermatology
17 January 2012
Number of times this article has been viewed

Greg J Goodman[1]
Michael B Halstead[2]
John D Rogers[2]
Daniela Borzillo[1]
Elizabeth Ryan[1]
Nick Riley[3]
John Wlodarczyk[3]

[1]Dermatology Institute of Victoria, South Yarra, Victoria, Australia; [2]Allergan Australia, Gordon, NSW, Australia; [3]John Wlodarczyk Consulting Services, New Lambton, NSW, Australia

Background: A software program called "HOYS" has been developed to depict various aspects and degrees of aging at 35 constituent subregions of seven distinct facial or exposed extrafacial regions. This program is underpinned by five-point photonumeric Likert scales characterizing skin surface and volume changes across five decades for each of the 35 subregions, and features an interactive skin-age assessment with a treatment-prioritization tool. In this study, the reliability and reproducibility of these scales was evaluated.

Methods: Eleven physicians and 19 non-physicians participated in this study. The five images from each of the 35 Likert scales in the HOYS program were shown on a total of 43 display boards, with selected subregions presented at rest or with movement, consistent with this program. Each image was randomly labeled between "A–E," corresponding to a range of skin ages by decade from 20–69 years. Each rater was asked to rank these images from youngest to oldest (or least to most severe deficit) for each scale and to repeat this exercise 2 hours later, with the intra- and inter-rater reliability evaluated. The raters were also asked to estimate the age of a single randomly allocated image on each scale for the purposes of internal validation.

Results: The overall inter-rater reliability of the raters was high at the first ranking session (weighted kappa: 0.78; 95% confidence intervals [95% CI]: 0.77–0.79) and this was confirmed when repeated 2 hours later (0.82; 95% CI: 0.81–0.83), with an intra-rater reliability of 0.76 (95% CI: 0.75–0.77). There was no significant difference in the physicians' and non-physicians' rankings. The raters also accurately estimated the actual age of the single randomly allocated image from each of the 43 stations (0.72; 95% CI: 0.70–0.74). A very similar pattern was observed when the ratings of a constituent of one of the seven regions, the perioral/lower face, were analyzed for expounding purposes.

Conclusion: The high reliability and reproducibility of the ranking in this validation study suggests that the five-point photonumeric Likert scales used in the HOYS program are an accurate depiction of age-related changes over five decades in the seven facial and extrafacial regions represented in this program, from the ages of 20–69 years.

Keywords: validation, inter-rater, intra-rater, reliability, age-related skin changes, HOYS, photonumeric scale

Correspondence: Greg J Goodman
Dermatology Institute of Victoria,
8-10 Howitt Street, South Yarra,
Victoria 3141, Australia
Tel +61 3 9826 4966
Email gg@div.net.au

Introduction

Age-related changes are not consistent in the facial or exposed extrafacial regions of women of similar age, nor are they symmetrical for the same woman. Many women have distinct areas (subregions) of the face, chest, neck, or hands, which may make them look older than they are chronologically. Conversely, they may have other sub-

© 2012 Goodman et al, publisher and licensee Dove Medical Press Ltd. This is an Open Access article which permits unrestricted noncommercial use, provided the original work is properly cited.

regions that are perceived to be more consistent with their actual age. This situation may be compounded by the absence of reliable methodologies for the evaluation of skin surface and volume changes associated with aging, potentially resulting in inconsistencies in diagnoses and indeed, in what esthetic treatments are administered. A new patient education software program, called "HOYS" for *Home of Younger Skin*, has been developed with the objective of creating an individualized and reproducible consultation, empowering patients to make informed treatment decisions for their facial and extrafacial correction/rejuvenation, in coalition with their health care practitioner.

The HOYS program has two potential independent audiences: the patient (and their esthetic physician) and the clinical investigator. For the patient, HOYS is specifically designed for the self-evaluation of their facial and environmentally exposed extrafacial features. As the patients complete the program, they will educate themselves on the age-related changes encountered by women from their twenties to their sixties through viewing a series of representative digital images. These images characterize age-related changes in skin surface and volume at a total of 35 defined areas (subregions) of the human face, at rest or during movement. Each of these 35 subregions is depicted by five images from a "typical" patient at the ages of 25, 35, 45, 55, or 65 years, reflecting five grades of aging or "severity" of the particular feature, presented in a Likert scale. The first program profile was designed for Caucasian females.

During the administration of HOYS, patients look at their image in a mirror while using a web-based application of the program, selecting the representative image that most closely matches their own appearance for the particular subregion under examination. The program is automated, thus allowing the patients to navigate through this assessment at their own pace, following the various prompts. At the conclusion of the program, the images and associated grades identified by the patients as closest to their own appearance will generate a report detailing the "skin age" for each of these seven regions, which are underpinned by the 35 subregions. These seven regions encompass the upper face, which is divided into two regions (forehead and temples, and the periorbital region); the midface; the perioral/lower face; and the extrafacial regions of neck, décolletage, and hands. A total skin age score is also generated, which is a composite of these seven regions. A treatment prioritization list is created by the HOYS program based on the degree of divergence of the age or "grade" assigned by the patient for each of the regions, relative to what is expected for a woman of the same chronologic age as defined by the program utilizing a proprietary algorithm. This forms the basis for a clinical treatment plan to be used by the patient's physician in a subsequent consultation with the patient.

As the HOYS program generates a total skin score, as well as the skin scores for the seven constitute regions detailed above, another potential use for HOYS is in the field of clinical research. By completing the HOYS-based assessment prior to treatment intervention and then repeating this assessment post-treatment, a quantitative measure of outcome can be captured. This could be employed as a means to assess comparative efficacy following the administration of different esthetic treatments.

The photonumeric Likert scales representing the 35 subregions employed in HOYS were created by an Australian dermatologist and primary author of this report (GG), who developed these scales and the overall program following an exhaustive review of over 5000 relevant images from his medical practice. The selection and associated ranking process used in the generation of these scales is therefore open to examination, and indeed confirmation, by the dermatologist's peers and by allied health care professionals working in esthetic medicine. This is to ensure that the scales employed in this program are in fact dependable and reproducible instruments, thereby validating the HOYS program. This was the principal objective of the study described in this report.

Methods

A group of 30 raters agreed to participate in this validation study. The raters consisted of eleven expert physicians (dermatologists, plastic surgeons, and cosmetic physicians) and 19 non-physicians (nurses and other nonmedical clinic staff), none of whom had previously been exposed to the program and received only a brief overview immediately prior to their participation in the study.

In preparation for the study, individual image stations were positioned in two large meeting rooms. Each station consisted of an A2-sized display board with five digital images (160 mm × 120 mm) from the photonumeric Likert scales for the 35 subregions employed in the actual HOYS program. For selected subregions, additional stations were utilized to represent supplementary aspects of the same area during facial animation (ie, smiling, frowning, squinting, raising eyebrows, pursing lips, or contracting neck muscles) to mimic the HOYS program, resulting in a total of 43 stations. (A list of the HOYS subregions represented on the 43 image stations is provided in Table 1). The images were positioned in random order on each display board, based on a block randomization of five, with the assigned labeling of A–E.

In one case, a display board had a total of ten images, representing juxtaposed lateral and frontal images of eyelashes for each of the five age ranges, with the same A–E labeling and block randomization applied.

Table 1 Seven geographical regions and 35 constitute subregions of the HOYS program[a]

Region	Subregions assessed
Forehead and temple	Forehead lines at rest
	Forehead lines on contraction
	Forehead sebaceous hyperplasia
	Glabella furrows at rest
	Glabella furrows on contraction
	Temples
Periorbital	Female brow position and shape
	Upper eyelids
	Crow's feet at rest
	Crow's feet on contraction
	Eyelashes (lateral/frontal)
	Infraorbital lines under the eyes at rest
	Infraorbital lines under the eyes on contraction
	Nasojugal folds and tear trough deformities (dark circles)
	Infraorbital fat pads
Midface and nose	Nasal droop
	Bunny scrunch or wolf lines at rest
	Bunny scrunch or wolf lines on contraction
	Facial wrinkling at rest
	Facial wrinkling on contraction
	Pores
	Facial indented scars
	Face telangiectasias (displayed on cheeks)
	Facial "age" spots
	Facial sunspots and other sun-induced lesions
	Face dyschromia (blotchy coloring)
	Facial shape
	General facial volume
	Malar (cheek) volume
Perioral/lower face	Nasolabial folds
	Upper lip atrophy
	Lip volume
	Upper lip wrinkles at rest
	Upper lip wrinkles on contraction
	Marionette lines and prejowl sulcus
	Jaw line
Neck	Neck bands at rest
	Neck bands on contraction
	Chin and neck loss of definition and redundancy
	Neck surface
Décolletage	Décolletage
Hands	Hands fullness and elasticity
	Hands skin quality

Note: [a]For selected subregions, additional stations were utilized to represent supplementary aspects of the same area during facial animation (ie, smiling, frowning, squinting, raising eyebrows, pursing lips, or contracting neck muscles) to mimic the HOYS program, resulting in a total of 43 stations.
Abbreviation: HOYS, Home of Younger Skin.

The raters were asked to sequentially rank the images at each station: (1) representing the youngest skin age and (5) the oldest. This was documented on a case report form (CRF) identified by the rater's ID (initials and date of birth). They were asked to complete this ranking exercise for all 43 stations, spending a maximum of 1 minute at each, and were not to discuss their rankings with their colleagues. Each rater then returned the completed CRF to the study coordinator (Procedure 1). Without being able to review their rankings recorded during Procedure 1, each rater was asked to repeat the exercise 2 hours later using a new CRF (Procedure 2). To evaluate internal consistency, upon completion of Procedure 2, raters were then asked to return to each station and record their best guess of the age represented by a particular image, which had also been designated through a block randomization and identified by a star symbol (Procedure 3).

For the calculation of inter-rater reliability, all available data pairs were used in the analysis. Agreement between the raters' ranking and true ranking were summarized by weighted kappa statistics and 95% confidence intervals (95% CI), with the former measuring agreement beyond chance, taking into account how close the rating was to the true response. The reproducibility of raters' responses was summarized at two time points (Procedures 1 and 2) and via agreement between the true age groups and the estimated age groups (Procedure 3). In addition, bubble charts were used to graphically present the range of responses for Procedures 1 and 2. The data from Procedure 3 were also summarized by kappa statistics and 95% CI, comparing actual age band to the estimated age band. Furthermore, to illustrate how the program utilizes the components (subregions) of the seven regions, one of these regions, the perioral/lower face, was selected, with the raters' ranking of the constituent subregions analyzed. This region comprises the nasolabial folds (upper lip atrophy, lip volume, upper lip wrinkles; at rest or on contraction, marionette lines, prejowl sulcus, and jaw line). The representative images of each of the aforementioned subregions, which were used in this validation study (albeit in random order), are correctly presented in Figure 1 (A–G), depicting youngest to oldest or "least to most severe."

Results

Inter-rater reliability assessment (all seven regions)

Procedure 1

A high degree of reliability in sequencing the images for all seven regions was confirmed when assessed for all 30 raters, with a weighted kappa of 0.78 (95% CI: 0.77–0.79). The eleven physicians (0.79; 95% CI: 0.78–0.81) and

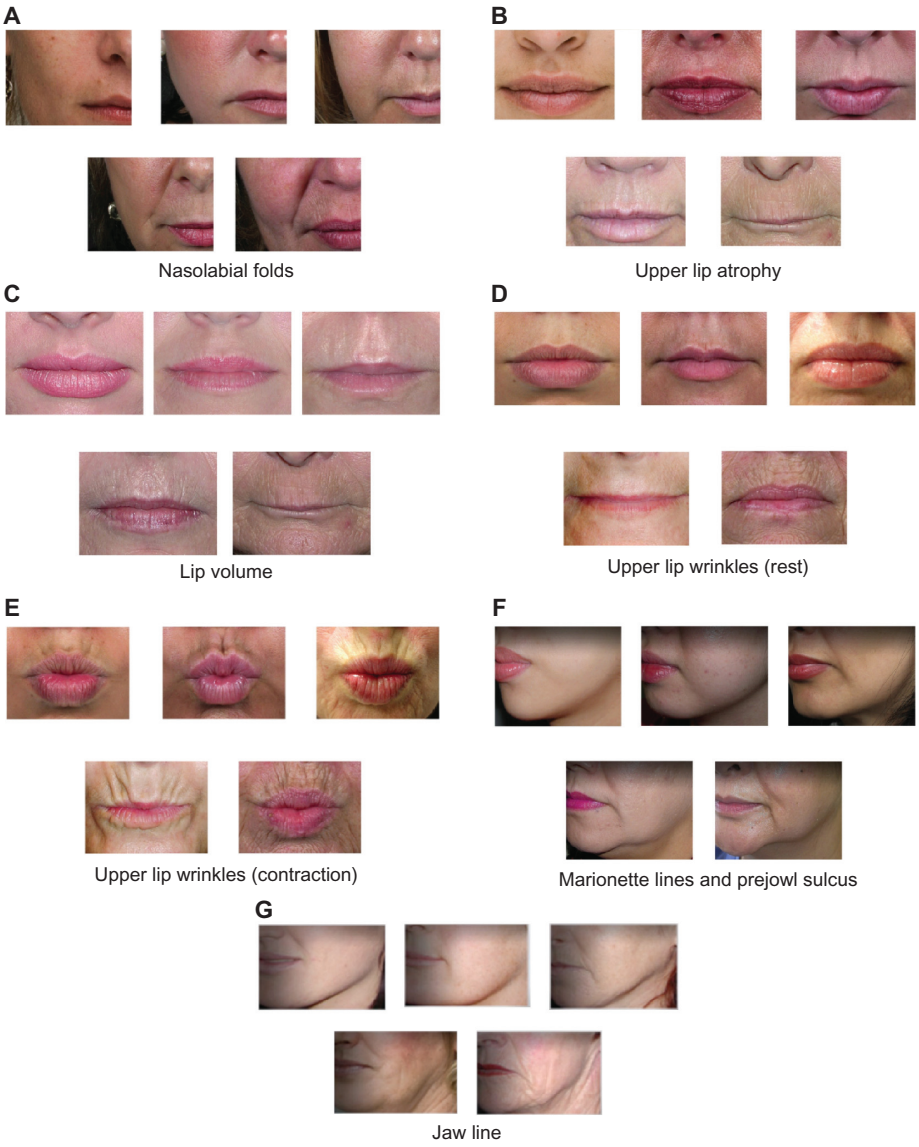

Figure 1 Grading (in correct order from youngest (1) to oldest (5) or "least to most severe").
Notes: (**A**) Nasolabial folds, Grade 1: Barely perceptible; Grade 2: Shallow, just perceptible; Grade 3: Moderately deep; Grade 4: Severe; Grade 5: Extremely overlapping. (**B**) Upper lip atrophy, Grade 1: No flattening; Grade 2: Mild flattening; Grade 3: Moderate flattening, mild wrinkling mainly due to volume loss; Grade 4: Moderate wrinkling, moderate lengthening of the distance between nose and lip border due to volume loss, some yellowing and sun damage; Grade 5: Severe wrinkling and wizened appearance, marked lengthening of the distance between nose and lip border due to volume loss. (**C**) Lip volume, Grade 1: Youthful full-bodied lips; Grade 2: Mild thinning of lips, minimal number of lip wrinkles breaking up the lip border, slight loss of definition of bow shape; Grade 3: Moderate thinning of lips, moderate number of lip wrinkles breaking up the lip border, moderate loss of definition of bow shape; Grade 4: Moderate to severe thinning of lips, severe number of lip wrinkles breaking up the lip border; moderate loss of definition of bow shape; Grade 5: Severe thinning of lips, severe number of lip wrinkles breaking up the lip border, severe loss of definition of bow shape. (**D** and **E**) Upper lip wrinkles at rest and on contraction, Grade 1: No wrinkles at rest or when pursing the lips; Grade 2: Mild wrinkles at rest and when pursing the lips; Grade 3: Mild wrinkles at rest, moderate when pursing the lips; Grade 4: Moderate wrinkles at rest, severe when pursing the lips; Grade 5: Moderate wrinkles at rest, severe when pursing the lips. (**F**) Marionette lines and prejowl sulcus, Grade 1: No marionette lines or jowl dip; Grade 2: Slight turn down at the corners of the mouth; Grade 3: Moderate marionette lines, mild jowl dip; Grade 4: Severe marionette lines extending towards the chin, moderate jowl dip; Grade 5: Severe marionette lines almost reaching the chin, severe jowl dip. (**G**) Jaw line, Grade 1: Tight jaw line; Grade 2: Softening of jaw line definition; Grade 3: Some blurring of jaw line and loosening of tissues with mild jowl formation; Grade 4: Indistinct jaw line with quite obvious jowls; Grade 5: Significant sagging eliminating jaw line definition; severe jowls.

19 nonphysicians (0.77; 95% CI: 0.76–0.79) ranked the images similarly (Table 2). Furthermore, when analyzed separately, the seven individual regions displayed a consistent weighted kappa of between 0.74 for the neck and 0.88 for the hands (Table 3). These data are also presented graphically in a bubble chart (Figure 2).

Procedure 2 (Procedure 1 repeated 2 hours later)

Following a 2-hour interval, Procedure 1 was repeated, with similar results achieved: a weighted kappa of 0.82 (95% CI: 0.81–0.83) for 28 raters (one physician and one nonphysician did not complete Procedure 2), with no differences in the ranking of the physicians (0.82; 95% CI: 0.80–0.84) and

Table 2 Inter- and intra-rater reliability (all seven regions)

All seven regions	Raters (n)	Observations (n)	Agreement (weighted kappa)	95% CI
Procedure 1				
All raters	30	6316	0.78	0.77–0.79
Physician	11	2240	0.79	0.78–0.81
Non-physician	19	4076	0.77	0.76–0.79
Procedure 2				
All raters	28	5977	0.82	0.81–0.83
Physician	10	2143	0.82	0.80–0.84
Non-physician	18	3834	0.82	0.81–0.83
Procedures 1 vs 2				
All raters	28	6273	0.76	0.75–0.77
Physician	10	2232	0.75	0.73–0.77
Non-physician	18	4041	0.76	0.75–0.78
Procedure 3				
All raters	28	1202	0.72	0.70–0.74

Abbreviations: CI, confidence interval; n, number.

nonphysicians (0.82; 95% CI: 0.81–0.83) (Table 2). Again, the agreement for the seven individual regions was high, with a weighted kappa ranging between 0.77 for the décolletage and 0.87 for the hands (Table 3). These data are also presented graphically in a bubble chart (Figure 3).

Intra-rater reliability assessment

When the reliability of each rater for all seven regions was analyzed for Procedures 1 versus 2, a high intra-rater consistency was confirmed: a weighted kappa for all raters of 0.76 (95% CI: 0.75–0.77), with similar results documented for physicians; 0.75 (95% CI: 0.73–0.77) and nonphysicians; 0.76 (95% CI: 0.75–0.78) (Table 2). The intra-rater agreement between Procedures 1 and 2 for the seven individual regions was again demonstrated, with a weighted kappa ranging between 0.66 for décolletage and 0.82 for neck (Table 3).

Procedure 3 (internal consistency: raters' estimation of actual age)

From the analysis of the age estimates of the raters for the randomly assigned photo at each display station, a satisfactory weighted kappa of 0.72 (95% CI: 0.70–0.74) was demonstrated, based on the ranking of all the raters (Table 2).

Inter- and intra-reliability assessment (perioral/lower face region constituents)

To assess for a consistency across the constituents of a particular region, the same analysis was conducted for the individual components (subregions) of the perioral/lower face region. A very consistent pattern of high inter- and intra-rater agreement was confirmed for these subregions, with a weighted kappa of 0.79 or above for the evaluation of inter-rater reliability during Procedures 1 and 2 (Table 4). The only exceptions to this were Upper lip wrinkling at rest in Procedure 1 (0.66; 95% CI: 0.58–0.74) and Jaw line in Procedures 1 (0.64; 95% CI: 0.56–0.72) and 2 (0.67; 95% CI: 0.59–0.75). The overall intra-rater reliability was also very high, with a weighted kappa of 0.75 or greater, with the exception of Lip volume (0.69; 95% CI: 0.61–0.77) and the Jaw line (0.66; 95% CI: 0.58–0.75) (Table 4).

Table 3 Inter- and intra-rater reliability assessment (all raters by the seven individual regions)

Region	Procedure 1		Procedure 2		Procedures 1 vs 2	
	Agreement (weighted kappa)	95% CI	Agreement (weighted kappa)	95% CI	Agreement (weighted kappa)	95% CI
Forehead and brow	0.80	0.77–0.82	0.83	0.81–0.86	0.76	0.73–0.79
Periorbital	0.78	0.75–0.80	0.79	0.77–0.82	0.73	0.70–0.76
Midfaces	0.76	0.74–0.78	0.81	0.79–0.83	0.77	0.75–0.79
Perioral/lower face	0.81	0.79–0.83	0.85	0.82–0.87	0.75	0.73–0.77
Neck	0.74	0.71–0.78	0.82	0.79–0.85	0.82	0.79–0.86
Décolletage	0.78	0.72–0.84	0.77	0.70–0.83	0.66	0.58–0.74
Hands	0.88	0.84–0.93	0.87	0.82–0.92	0.71	0.66–0.77

Abbreviation: CI, confidence interval.

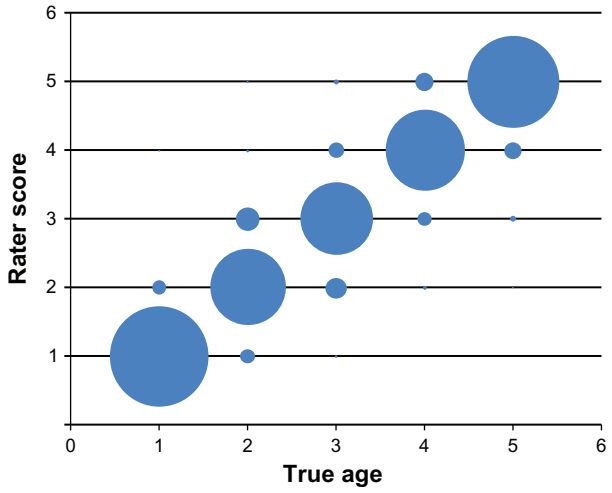

Figure 2 Rater score versus true age during Procedure 1 (all raters, all regions).

Raters' estimation of age (perioral/lower face region constituents)

A scatter graph showing the individual age estimations of the raters, plotted against the correct age for the individual components of the sample perioral/lower face region, is also presented for the constituents of this region (Figure 4). It also highlights the good level of accuracy and agreement between raters.

Discussion

Inherently, when we meet people for the first time we judge their age, or when we greet someone whom we haven't seen for a long time we assess how they have aged since we last met, and indeed, whether the person looks "young or old" for their age. How do we do this? It is probable that we closely analyze the features of another person's face subconsciously, perhaps

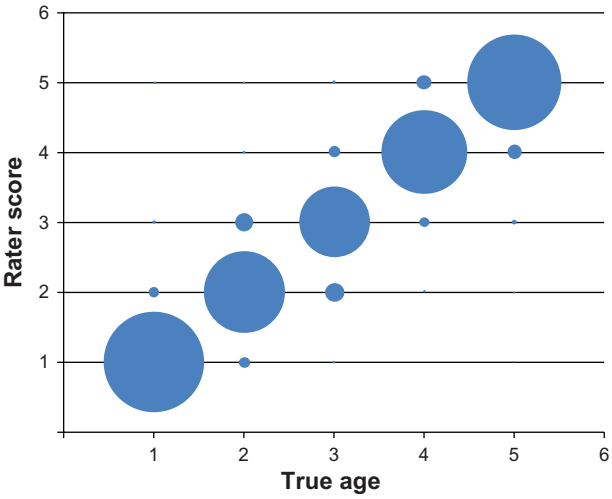

Figure 3 Rater score versus true age during Procedure 2 (all raters, all regions).

Table 4 Inter- and intra-rater reliability assessment (all raters for the perioral/lower face region)

Perioral/lower face subregion	Observations (n)	Procedure 1		Observations (n)	Procedure 2		Observations (n)	Procedures 1 vs 2	
		Agreement (weighted kappa)	95% CI		Agreement (weighted kappa)	95% CI		Agreement (weighted kappa)	95% CI
Nasolabial folds	145	0.80	0.74–0.86	140	0.88	0.82–0.93	145	0.77	0.70–0.83
Upper lip atrophy	145	0.83	0.77–0.88	140	0.89	0.85–0.94	145	0.86	0.81–0.91
Lip volume	145	0.97	0.94–0.99	140	0.94	0.90–0.97	140	0.69	0.61–0.77
Upper lip wrinkling (at rest)	145	0.66	0.58–0.74	138	0.79	0.72–0.85	138	0.79	0.72–0.85
Upper lip wrinkling (contracted)	145	0.85	0.81–0.90	140	0.84	0.79–0.89	140	0.75	0.69–0.81
Marionette and prejowl sulcus	143	0.91	0.87–0.95	140	0.92	0.87, 0.98	140	0.85	0.79–0.91
Jaw line	143	0.64	0.56–0.72	140	0.67	0.59–0.75	143	0.66	0.58–0.75

Abbreviations: CI, confidence interval; n, number.

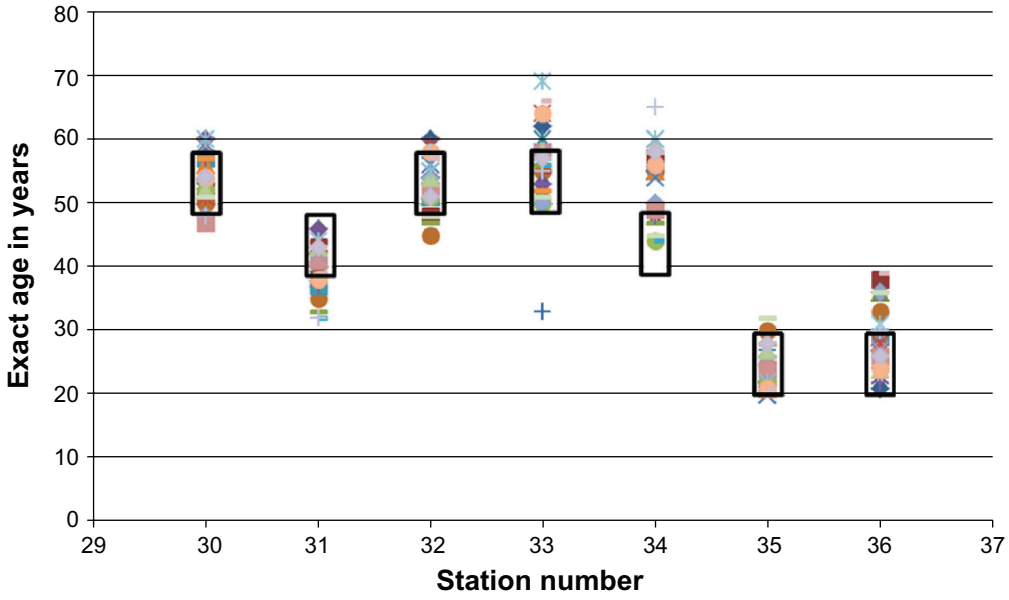

Figure 4 Constituents of the perioral/lower face region; exact age estimations during Procedure 3 (all raters, stations 30–36).
Notes: Stations: 30 Nasolabial folds; 31 Upper lip atrophy; 32 Upper lip wrinkles at rest; 33 upper lip wrinkles on contraction; 34 Lip volume; 35 Marionette and prejowl sulcus; 36 Jaw line. The rectangles represent the actual age decade for the single randomized photo at each station with data points being the rater's estimate of the age from the subregion photographic representation.

comparing what we see against a hardwired understanding of how particular features of a person should present for any given age. The different visual clues one obtains from facial and exposed extrafacial regions of the other person may quickly provide us with the necessary information to process.

It is well established that our genetics meshed with the environment to which we are exposed ultimately determines whether our current appearance or "esthetic age" is consistent with our chronological age. That is, "are we aging better or worse, relative to others of the same vintage?" Individuals of certain ethnicities, for example, may enjoy a comparative delay in the formation of lines and wrinkles owing to a greater deposition of collagen and elastin in their skin. However, this outcome might still be significantly altered by the environment, in addition to the lifestyle of the individual, with the potential for skin changes to occur nonuniformly. Those who smoke regularly, for example, may develop premature perioral lines due to chronic movement of the orbicular oris muscle over many years. Other individuals who have continuously squinted due to chronic exposure to bright sunlight may present with wrinkles in the lateral canthal region, which are consistent with someone several years older than their chronological age.

To address the above issues, the HOYS program was designed to facilitate a detailed self-assessment of 35 subregions representing the constituents of seven regions across the face, the neck, the décolletage, and the hands. Using a proprietary algorithm, differences from one region to another can be quickly identified, with a composite (total) age score, as well as the score for each of these seven regions recorded, thus quantifying any divergence between a patient's esthetic and chronological age. By following this methodology, those areas that appear to have aged disproportionately (that is, accentuating the greatest imbalance in one's overall features) can be quickly determined. A specific treatment protocol can subsequently be employed to achieve a harmonious outcome; a departure from the practice of assessing specific features of a patient's face in isolation. Since these scores are generated during the patient's own self-evaluation, a greater understanding of the aging process at each region, and indeed, at the constituent subregions is likely to be obtained. This improved knowledge may facilitate a more active participation of the patient in any subsequent physician consultation with regard to treatment options. This might be focused on the removal of pigmentation, the correction or minimization of wrinkles and skin folds, or the restoration of lost volume or structure, amongst the myriad of rejuvenation and/or enhancement strategies available to the esthetic physician.

Many other grading scales have been described previously to rank specific facial and/or extrafacial subregions.[1–13] In addition, several facial recognition techniques have been perfected over the years to analyze facial features and may also be programmed to document changes as a person ages.[14–16]

The HOYS program differs from these aforementioned technologies by:

- Providing the patient with education on age-related changes to skin surface and volume, which is individualized. This information thereby empowers the patient to actively participate in any subsequent discussions with their esthetic physician about the diagnosis of any deficits and the rationale for the treatment recommendations. (This discussion might otherwise be a predominantly physician-driven exercise due to a general lack of understanding of the above on the part of the patient).
- Allowing variances in subregions to be isolated and highlighted, thereby providing the basis for a transparent treatment plan for the patient which can be prioritized to address the greatest variance.
- Permitting repeat examination for any treated area to quantify any improvement or otherwise at the subregional or regional level, which can be subsequently used for patient consultation in the clinic, or indeed for research purposes to assess outcome for one or more esthetic treatments.

In this validation study, a very high agreement was demonstrated for the raters for the seven facial and extrafacial regions (Tables 2 and 3), highlighting the intuitive nature of the five-point photonumeric scales that are the backbone of the HOYS program. This was also the case at the subregion level when the rankings of the constituents of the perioral/lower face were analyzed (Table 4). The only exception to this was the ranking of the Upper lip wrinkling at rest scale in Procedure 1, although an acceptable weighted kappa of 0.66 (95% CI: 0.58–0.74) was still returned and agreement was high (0.79; 95% CI: 0.72–0.85) for the same scale when it was ranked in Procedure 2 (Table 4). Moreover, the related scale; Upper lip wrinkling on contracture, had a weight kappa of 0.85 (95% CI: 0.81–0.90) and 0.84 (95% CI: 0.79–0.89) for Procedures 1 and 2, respectively (Table 4). Notably, when these animated images are paired with the equivalent images at rest, as would be the case when the "live" HOYS-based evaluation is conducted by patients, little difficulty in grading would be expected. The ranking of the Jaw line was also not as reliable as for the other subregions. However, it was still quite acceptable with a weighted kappa of 0.64 (95% CI: 0.56–0.72) in Procedure 1, rising to 0.67 (95% CI: 0.59–0.75) in Procedure 2 (Table 4). Intra-rater reliability was also high for the perioral/lower face subregions, with agreement at 0.75 or above, with the exception of Lip volume and Jaw line, although both subregions still had a very respectable weighted kappa of 0.69 (95% CI: 0.61–0.77) and 0.66 (95% CI: 0.58–0.75), respectively (Table 4).

It is evident from the bubble charts that the agreement between raters during both Procedures 1 and 2 (Figures 2 and 3) was lower for the images of subjects in the mid decades of life, relative to those representing the age groups of 20–29 and 60–69 years. Nevertheless, a weighted kappa of 0.70 or greater was still calculated for these mid-age groups. Explanations for this phenomenon will to some degree be found in the nature of these or similar scales, with the youngest or least affected and the oldest or most affected generally easier to rate since there is no alternate choice below or above these grades, respectively. This may also reflect the relatively slower, more subtle aging processes in Caucasian females between the ages of 30–45 years, which generally hastens thereafter with the start of menopause.

There should be a clear demarcation between the grades on the five-point Likert scales used for each subregion in the HOYS program. The results from Procedure 3 (Table 2; Figure 4) show a high degree of accuracy and agreement amongst the raters for age estimation, based on only a single randomized image of each of the representative subregions in this program. For many of these subregions only a small segment of the facial or extrafacial area in question is displayed in the relevant image (see Figure 1A–G). This may be consistent with our innate ability to calculate another person's age from only subtle visual clues, as described above.

Conclusion

In summary, the high inter- and intra-rater reliability of the rankings, as documented by both physicians and nonphysicians in this validation study, suggests that the specific five-point photonumeric scales that underpin the HOYS education program are an accurate representation of age-related changes likely to be observed in the 35 constitute subregions of seven distinct facial and exposed extrafacial regions of individuals, ranging between 20–69 years. As a consequence, this interactive software program, designed for self-evaluation by patients, may serve as a useful scale in esthetic practices to evaluate age-related skin changes. It may also be an effective research tool to quantitatively evaluate the therapeutic effects of different esthetic treatments. Importantly, due to the individualized outputs resulting from their own self-evaluation, patients completing this program may perceive greater empowerment in any subsequent consultation with their physician on their treatment options through an improved understanding of the rationale for these treatments, and indeed, how these treatments might be

prioritized to best enhance their overall appearance. This may result in logic-based treatment decisions by the patient, as these judgments may no longer be based on a particular bias that a patient or a physician may have. Such an approach may result in better treatment outcomes for patients, as evidenced by improved patient satisfaction and greater overall fiscal investment in enhancing his or her appearance. Consistent with this, a high level of patient satisfaction with the HOYS program was confirmed in a prospective, randomized, controlled study, in which esthetically orientated patients compared the utility of HOYS to a standard patient education program on age-related skin changes currently employed in many esthetic clinics throughout Australia.[17]

Disclosure

GG developed the HOYS program. MH and JR are employees of Allergan Australia, which sponsored the study. JW and NR are employed by John Wlodarczyk Consulting Services, which conducted the data management and the analysis. ER and DB are employees of HOYS Pty Limited. None of the authors participated as raters in this validation study.

References

1. Carruthers A, Carruthers J, Hardas B, et al. A validated hand grading scale. *Dermatol Surg*. 2008;34 Suppl 2:S179–S183.
2. Carruthers A, Carruthers J, Hardas B, et al. A validated grading scale for marionette lines. *Dermatol Surg*. 2008;34 Suppl 2:S167–S172.
3. Carruthers A, Carruthers J, Hardas B, et al. A validated grading scale for forehead lines. *Dermatol Surg*. 2008;34 Suppl 2:S155–S160.
4. Carruthers A, Carruthers J, Hardas B, et al. A validated lip fullness grading scale. *Dermatol Surg*. 2008;34 Suppl 2:S161–S166.
5. Carruthers A, Carruthers J, Hardas B, et al. A validated grading scale for crow's feet. *Dermatol Surg*. 2008;34 Suppl 2:S173–S178.
6. Carruthers A, Carruthers J, Hardas B, et al. A validated brow positioning grading scale. *Dermatol Surg*. 2008;34 Suppl 2:S150–S154.
7. Buchner L, Vamvakias G, Rom D. Validation of a photonumeric wrinkle assessment scale for assessing nasolabial fold wrinkles. *Plast Reconstr Surg*. 2010;126(2):596–601.
8. Kim EJ, Reeck JB, Maas CS. A validated rating scale for hyperkinetic facial lines. *Arch Facial Plast Surg*. 2004;6(4):253–256.
9. Day DJ, Littler CM, Swift RW, et al. The wrinkle severity rating scale: a validation study. *Am J Clin Dermatol*. 2004;5(1):49–52.
10. Rossi AB, Nkengne A, Stamatas G, et al. Development and validation of a photonumeric grading scale for assessing lip volume and thickness. *J Eur Acad Dermatol Venereol*. 2011;25(5):523–531.
11. Valet F, Ezzedine K, Malvy D, et al. Assessing the reliability of four severity scales depicting skin ageing features. *Br J Dermatol*. 2009;161(1):153–158.
12. Hund T, Ascher B, Rzany B. Smile Study Group. Reproducibility of two four-point clinical severity scores for lateral canthal lines (crow's feet). *Dermatol Surg*. 2006;32(10):1256–1260.
13. Goodman G. Botulinum toxin for the correction of hyperkinetic facial lines. *Australas J Dermatol*. 1998;39(3):158–163.
14. Lanitis A, Draganova C, Christodoulou C. Comparing different classifiers for automatic age estimation. *IEEE Trans Syst Man Cybern B Cybern*. 2004;34(1):621–628.
15. Geng X, Zhou ZH, Smith-Miles K. Automatic age estimation based on facial aging patterns. *IEEE Trans Pattern Anal Mach Intell*. 2007;29(12):2234–2240.
16. Park SW, Savvides M. Individual Kernel Tensor-subspaces for robust face recognition: a computationally efficient tensor framework without requiring mode factorization. *IEEE Trans Syst Man Cybern B Cybern*. 2007;37(5):1156–1166.
17. Williams LM, Alderman JE, Cussell G, et al. Patient's self-evaluation of two education programs for age-related skin changes in the face: a prospective, randomized, controlled study. *Clin Cosmet Investiq Derm*. 2011;4:149–159.

Clinical, Cosmetic and Investigational Dermatology Dovepress

Publish your work in this journal

Clinical, Cosmetic and Investigational Dermatology is an international, peer-reviewed, open access, online journal that focuses on the latest clinical and experimental research in all aspects of skin disease and cosmetic interventions. All areas of dermatology will be covered; contributions will be welcomed from all clinicians and basic science researchers globally. This journal is indexed on CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: http://www.dovepress.com/clinical-cosmetic-and-investigational-dermatology-journal