Measuring skin aging using optical coherence tomography in vivo: a validation study

Carina Trojahn
Gabor Dobos
Claudia Richter
Ulrike Blume-Peytavi
Jan Kottner
Measuring skin aging using optical coherence tomography in vivo: a validation study

Carina Trojahn, Gabor Dobos, Claudia Richter, Ulrike Blume-Peytavi, and Jan Kottner*
Charité-Universitätsmedizin Berlin, Clinical Research Center for Hair and Skin Science, Department of Dermatology and Allergy, Charitéplatz 1, Berlin 10117, Germany

Abstract. Dermal and epidermal structures in human skin change during intrinsic and extrinsic aging. Epidermal thickness is one of the most often reported parameters for the assessment of skin aging in cross-sectional images captured by optical coherence tomography (OCT). We aimed to identify further parameters for the non-invasive measurement of skin aging of sun-exposed and sun-protected areas utilizing OCT. Based on a literature review, seven parameters were inductively developed. Three independent raters assessed these parameters using four-point scales on images of female subjects of two age groups. All items could be detected and quantified in our sample. Interrater agreement ranged between 25.0% and 83.3%. The item scores “stratum corneum reflectivity,” “upper dermal reflectivity,” and “dermoeipidermal contrast” showed significant differences between age groups on the volar and dorsal forearm indicating that they were best able to measure changes during skin aging. “Surface unevenness” was associated with the skin roughness parameters, $R_s$ and $R_{max}$, on the inner upper arm and volar forearm supporting the criterion validity of this parameter on sun-protected skin areas. Based on the interrater agreement and the ability to differentiate between age groups, these four parameters are being considered as the best candidates for measuring skin aging in OCT images. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.20.4.045003]

Keywords: optical coherence tomography; clinical assessment; skin microtopography; skin aging.

1 Introduction

Skin aging is the continuous change of the skin’s structure and functional capacity. In addition to the “normal” course of intrinsic or chronological aging, environmental (e.g., sun light) and life style (e.g., smoking) factors contribute to the so-called extrinsic skin aging. Skin areas that are predominantly exposed to ultra violet radiation (e.g., face, hands, and dorsal aspect of the forearm) are prone to develop various signs of extrinsic aging. These include deep wrinkles, loss of elasticity, and changes in skin pigmentation. Histologically, extrinsically aged skin shows degenerated collagen and elastic fibers in the upper and middermis, which is referred to as solar elastosis.1–3

Facial appearance contributes to the perceived age of individuals and is an indicator of overall health.4,5 Thus, there is growing interest in skin aging research and so-called antiaging strategies. The evaluation of cosmetic products and rejuvenation procedures is often based on histological assessments that provide knowledge of the morphological and molecular characteristics of the skin.6,7 However, the invasive procedure to obtain biopsies contains disadvantages, such as tissue destruction, pain and an increased risk for infections.

Optical coherence tomography (OCT) allows a noninvasive and rapid cross-sectional visualization of the skin. It is a widely applied imaging method in many fields of dermatology,8,9 for instance, to follow up wound healing, to diagnose skin cancer, or to assess the effectiveness of skin care products.10–13 To date, only a few studies have used OCT to investigate skin aging. The most often reported parameter in skin aging research measured in OCT images is epidermal thickness (ET). ET has been often observed to decrease with age on intrinsically and extrinsically aged skin areas.14–16 However, OCT images obviously provide much more information than ET only. Therefore, the idea emerged, whether there are other traits or criteria that might be used to characterize signs of skin aging in OCT images. The aim of this study was to develop and to evaluate possible additional parameters to measure skin aging using OCT.

2 Methods

2.1 Study Design and Participants

An exploratory validation study was conducted at the Clinical Research Center for Hair and Skin Science, Department of Dermatology and Allergy, Charité-Universitätsmedizin Berlin, Germany, between January and April 2014. Caucasian women of two age groups with healthy skin were recruited. Written informed consent was obtained from all subjects before participation. They were instructed not to use any cosmetic products on their arms at least 12 h prior to the measurements.

2.2 OCT Measurements

The OCT system Telesto (Thorlabs, Lübeck, Germany) was used in this study for in vivo evaluation of skin aging. This measurement method is based on the detection of the optical path length differences of reflected and backscattered light of a sample and a reference beam (interferometry). The OCT system contains a low-coherence near infrared (about 1300 nm)
light source and a spectrometer that detects the phase delay of the wavelengths. The hand-held OCT probe was positioned directly onto the skin. Two-dimensional images were acquired with a scan length of around 10 mm, a lateral resolution of up to 8 μm, and a maximum penetration depth of around 1.5 mm.

ET was measured using the software ImageJ and a plug-in developed by Thorlabs. ET measurements were performed on three predefined measurement sites (center of the image and two lateral sites with a fixed distance to the center) by height-width titration, which is the length of a vertical line against the image per se. The average ET was calculated from these three measurements. All ET measurements were made by two investigators blinded to the participants’ ages and skin areas.

2.3 OCT Items and Score Development

In order to develop possible detectable characteristics of skin aging, a literature review was conducted in Medline. Studies that applied different skin imaging techniques, such as sonography, histology, and laser scanning microscopy, as well as studies reporting features of skin morphology, wound healing, and pressure ulcer development were reviewed. Based on these reports, seven possible characteristics were developed inductively and iteratively by the authors. The items were discussed with a physician and a researcher, both experienced in OCT imaging, and further adapted. The preliminarily determined items were defined as follows: (1) “stratum corneum (SC) reflectivity,” (2) “epidermal reflectivity,” (3) “upper dermal reflectivity,” and (4) “lower dermal reflectivity.” These four reflectivity items describe the intensity of the signal detected by the camera backscattered from the respective tissues.

In a next step, OCT images taken from the right inner upper arm, right volar forearm, and right dorsal forearm of each participant were evaluated by three raters experienced in OCT imaging (two blinded to the participants’ ages and one unblinded author) independently from each other. The raters were trained by the first author on item scale definitions using example images not included in the current analysis. Each item was graded using a four-point scale (0, absent; 1, low; 2, moderate; 3, high), except “vessel density,” which was graded using a four-point scale ranging from 0 = no vessels to 3 = many vessels. This four category classification was applied to avoid a neutral position and to ensure differentiation.

2.4 Criterion Validation of the Item “Surface Unevenness”

Criterion validity of the item “surface unevenness” was explored by correlating the score with instrumented measures of the skin microtopography of the identical skin areas. The Visioscan® VC 98 camera and the corresponding software SELS 2000 (both Courage & Khazaka Electronic GmbH, Cologne, Germany) were used to measure the DIN/ISO parameters $R_s$ (arithmetic mean roughness from five consecutive sampling lengths) and $R_{\text{max}}$ (maximal roughness). Both parameters were chosen because they exhibited the highest reliability in comparable age groups. Roughness measurements were conducted in a controlled room temperature of 22 ± 2°C and 50% ± 10% relative humidity and after the participants’ acclimatization for at least 30 min.

2.5 Statistical Analysis

Demographic characteristics of the sample were described using means, standard deviations (SD), ranges, and frequencies. OCT item scores were analyzed descriptively using medians and interquartile ranges. Interrater agreement ($p_o$) of item scores between the three raters was calculated. This expresses the proportion in which the three raters achieved identical results per OCT image. Construct validity was evaluated using Mann–Whitney $U$ tests to compare medians of the item scores between the two age groups per skin area. ET and skin surface roughness were presented as means and SDs. Group comparisons were performed using student’s $t$ test for independent samples. Criterion validity was tested using Spearman’s correlation coefficients ($r_s$) between the OCT item score “surface unevenness” and skin roughness measurements $R_s$ and $R_{\text{max}}$. A $p$-value of <0.05 (two-sided) was considered statistically significant. Statistical analysis was performed using IBM SPSS Statistics, Version 22.0 (IBM Corp., Armonk, New York).

3 Results

3.1 Sample Characteristics

Demographic characteristics of the study sample are summarized in Table 1. Sixteen female subjects were included in the validation study, eight subjects in the young group [mean age 33.5 (SD 2.1) years] and eight subjects in the old group [mean age 76.6 (SD 1.9) years]. Mean body mass indices were 21.8 (SD 2.0) kg/m² and 26.2 (SD 2.5) kg/m². Equal distributions of Fitzpatrick phototypes II and III were observed in both groups.

3.2 Epidermal Thickness

Mean ET ranged from 64.8 μm on the dorsal forearm to 78.1 μm on the inner upper arm of the young group (Table 2). Differences between age groups were not statistically significant.

<table>
<thead>
<tr>
<th>Table 1 Sample characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Age in years Mean (SD)</td>
</tr>
<tr>
<td>range</td>
</tr>
<tr>
<td>BMI in kg/m² Mean (SD)</td>
</tr>
<tr>
<td>range</td>
</tr>
<tr>
<td>Phototype, n</td>
</tr>
</tbody>
</table>

Note: BMI, body mass index and SD, standard deviation.
Table 2  Epidermal thickness (μm) measured in optical coherence tomography (OCT) images.

<table>
<thead>
<tr>
<th></th>
<th>Young group</th>
<th>Old group</th>
<th>p value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Inner upper arm</td>
<td>78.1 (17.0)</td>
<td>69.8 (12.8)</td>
<td>0.284</td>
</tr>
<tr>
<td>Volar forearm</td>
<td>65.1 (8.9)</td>
<td>67.9 (10.5)</td>
<td>0.581</td>
</tr>
<tr>
<td>Dorsal forearm</td>
<td>64.8 (12.0)</td>
<td>70.4 (18.9)</td>
<td>0.489</td>
</tr>
</tbody>
</table>

Note: SD, standard deviation.
<sup>a</sup>Student’s t test.

### 3.3 OCT Item Scores

All seven items were detected and could be quantified in our sample. Figure 1 shows the OCT images representing the lowest (left column) and highest (right column) mean score per item.

#### 3.3.1 Construct validation: skin aging

The grading results of the OCT images are shown in Table 3. Median scores ranged from 0.8 for the item “surface unevenness” to 2.5 for the item “upper dermal reflectivity.” No floor and ceiling effects were observed.

Median “SC reflectivity” was higher in older subjects on all three skin areas with differences being statistically significant for the forearm sites. “Upper dermal reflectivity” was significantly lower in the old group on both forearm areas. “Dermoepidermal contrast” and “vessel density” were higher in the old group on the forearm and comparable on the inner upper arm. “Epidermal reflectivity” and “lower dermal reflectivity” were reduced in the old group without reaching statistical significance. “Surface unevenness” increased with age on the inner upper arm and volar forearm and decreased on the dorsal forearm.

Differences between the skin areas were observed for “dermoepidermal contrast,” “vessel density,” and “surface unevenness.” Other item scores like “lower dermal reflectivity” and “upper dermal reflectivity” showed minor differences between the skin areas.

#### 3.3.2 Interrater agreement

Interrater agreement is presented in Table 3. The highest proportions of agreement were obtained for the items “SC reflectivity,” “upper dermal reflectivity,” “dermoepidermal contrast,” and “surface unevenness” (p<sub>α</sub> = 83.3%). The lowest proportion of agreement were observed for the item “epidermal reflectivity” (p<sub>α</sub> = 25.0%).

#### 3.3.3 Criterion validation: skin microtopography

The mean surface roughness parameters of the two age groups are presented in Table 4. R<sub>s</sub> ranged from 38.2 μm on the volar forearm in the young group to 55.0 μm on the dorsal forearm in the old group. R<sub>max</sub> ranged from 48.3 μm on the volar forearm of the young group to 74.1 μm on the dorsal forearm of the old group. SDs were always higher in the aged group. Differences between the age groups were statistically significant for the inner upper arm and volar forearm. In both age groups, surface roughness was highest on the dorsal forearm. Spearman correlation coefficients of the item scores of “surface unevenness” were highest with R<sub>max</sub> on the inner upper arm and on the volar forearm (r<sub>S</sub> = 0.442 and r<sub>S</sub> = 0.434). The lowest correlation coefficients were observed for R<sub>s</sub> and R<sub>max</sub> on the dorsal forearm (r<sub>S</sub> = 0.101 and r<sub>S</sub> = 0.151).

### 4 Discussion

This study introduces additional characteristics besides ET to measure skin aging in OCT images. In a sample of young and aged women, different sun-exposed and sun-protected skin areas were measured. Based on the literature and our own experience with OCT imaging, we inductively developed seven potentially observable skin characteristics.

The most common parameter measured in OCT images is ET, for which good correspondence with confocal laser scanning microscopy has been reported. The ET estimates obtained in our study were comparable with previously reported results. We were able to detect a slight thinning of the epidermis with age on the inner upper arm that confirms the results of previous studies. Nevertheless, ET thinning tends to occur considerably only up to the 30% and seems to be attenuated by the effects of photoaging.

All newly proposed characteristics were detected and could be quantified in our sample. The items had an overall good distribution of the scores and showed no floor and ceiling effects. This indicates that the item scores were suitable to discriminate between different skin characteristics.

Interrater agreement was high for the items “SC reflectivity,” “upper dermal reflectivity,” “dermoepidermal contrast,” and “surface unevenness,” indicating that these items were easier to grade leading to similar rating results. By contrast, agreement among the items “epidermal reflectivity” and “lower dermal reflectivity” was lower indicating that these items might be inadequate for assessing skin aging.

Study results demonstrated that there are age- and site-dependent differences regarding the item scores. We have chosen the construct “skin aging” to evaluate the validity of the rating results. In intrinsically aged skin, the SC turnover rate is much slower and corneocytes are larger. Moreover, older subjects have a reduced amount of intercellular lipids. These age-dependent changes may lead to alterations in the optical properties of the SC by affecting the penetration of light through the SC, which might be a possible explanation for the higher “SC reflectivity” on both forearm sites of the old group. Whether the intensive surface reflection is really caused by the structure of the SC or by slightly different refractive indices between air and skin is ambiguous. However, we observed an obvious difference between young and older skin.

Significant differences between the age groups were also observed on both forearm sites for “upper dermal reflectivity” and “dermoepidermal contrast,” which assumes that these parameters in OCT images do most clearly change with age. The decreased “upper dermal reflectivity” which was pronounced on both forearm sites likely reflects the deposition of disorganized and highly backscattering elastin from the papillary dermis to the reticular dermis and the loss of collagen content resulting in a signal attenuation of the upper dermis. In close relation to less “upper dermal reflectivity” on both forearm sites is the substantial increase in “dermoepidermal contrast.” The more the reflectivity of the upper dermis decreases the higher is the ability to distinguish between epidermis and dermis.
Fig. 1 Optical coherence tomography images selected among the 16 participants representing the lowest and highest mean scores per item and image. Mean scores are the averages of the scores of three raters. Scale bar 500 μm.
“Lower dermal reflectivity” seemed to be neither related to age nor to sun exposure. This finding is likely to be explained by the limited penetration depth and lateral resolution of the OCT signal leading to an increased signal noise and a diffuse appearance of the deeper dermis.

We expected to observe age- and site-dependent differences in “vessel density” because horizontal cutaneous vessels have been shown to be increased in the elderly.34 In addition, there is evidence that photodamaged skin exhibits reduced numbers of dermal vessels compared to intrinsically aged skin.35 On
the other hand, an increased scattering coefficient of the dermis can lead to the reduction of dermal reflectivity thus affecting the ability to visualize blood vessels. However, an age- and site-dependent change in “vessel density” could be shown in our sample supporting the construct validity of this item.

We observed an increase in “surface unevenness” in older subjects indicating a rougher and more folded skin surface in this age group. Epidermal changes and dermal degeneration processes during aging are considered to promote increased skin roughness. We could confirm this finding by Visioscan® measurements. and were, therefore, chosen as reference standards for criterion validation in this study. Mean and estimates were similar to previous findings in young and old women. All measured roughness estimates increased with age in this study, which is in line with previously reported studies. Validity of the item “surface unevenness” was demonstrated by correlation with both roughness estimates. The item scores were associated with and on the inner upper arm and volar forearm, supporting that these estimates measure a similar concept on these sun-protected skin areas. In conjunction with a slightly lower surface unevenness score on the photoaged dorsal forearm skin in the old group, our results support the hypothesis that chronic UV exposure reduces the effect of intrinsic aging on this skin area, indicating that sun damage counteracts increasing roughness during aging. Furthermore, the limited range of and values at the dorsal forearm attenuates observed validity coefficients thus explaining the low correlation.

4.1 Limitations
This explorative study was conducted with a small number of subjects to obtain first indicators about possible characteristics of OCT images that might be related to changes due to skin aging. Generalizability of the results is limited and larger confirmatory studies are needed. In order to reduce a possible bias due to gender and ethnicity, we included female Caucasians only. Finally, comparability with other OCT devices might be limited.

4.2 Conclusion
This study describes new quantifiable skin characteristics of OCT images besides ET. The items “SC reflectivity,” “upper dermal reflectivity,” “dermoeipidermal contrast,” and “surface unevenness” have shown sufficient interrater agreement and the ability to differentiate between age groups, thus being considered as the best candidates for measuring skin aging on OCT images.

Acknowledgments
The study was supported by the Clinical Research Center for Hair and Skin Science.

References

Table 4 Skin surface roughness parameters $R_z$ ($\mu m$) and $R_{\text{max}}$ ($\mu m$) measured using Visioscan® VC98.

<table>
<thead>
<tr>
<th></th>
<th>Young Mean (SD)</th>
<th>Old Mean (SD)</th>
<th>$p$ value</th>
<th>$r_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_z$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inner upper arm</td>
<td>39.1 (6.2)</td>
<td>53.3 (9.5)</td>
<td>0.003</td>
<td>0.412</td>
</tr>
<tr>
<td>Volar forearm</td>
<td>38.2 (4.3)</td>
<td>46.5 (7.1)</td>
<td>0.013</td>
<td>0.351</td>
</tr>
<tr>
<td>Dorsal forearm</td>
<td>48.6 (3.3)</td>
<td>55.0 (11.2)</td>
<td>0.148</td>
<td>0.101</td>
</tr>
<tr>
<td>$R_{\text{max}}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inner upper arm</td>
<td>50.0 (7.6)</td>
<td>69.0 (11.1)</td>
<td>0.001</td>
<td>0.442</td>
</tr>
<tr>
<td>Volar forearm</td>
<td>48.3 (5.0)</td>
<td>60.5 (9.0)</td>
<td>0.005</td>
<td>0.434</td>
</tr>
<tr>
<td>Dorsal forearm</td>
<td>65.0 (5.6)</td>
<td>74.1 (14.5)</td>
<td>0.122</td>
<td>0.151</td>
</tr>
</tbody>
</table>

Note: SD, standard deviation and $r_s$, Spearman’s correlation coefficient. Student’s $t$ test.