Testing the Reliability of Optical Coherence Tomography to Measure Epidermal Thickness and Distinguish Volar and Nonvolar Skin

Molly E. Baumann[1,2], Nina Rossa Haddad[3], Alyssa Salazar[2,4], W. Lee Childers[1,2,5], Shawn Farrokhi[1,6], Neil B. Goldstein[1,7], Brad D. Hendershot[1,5,7], Lisa Reider[8], Richard E. Thompson[8], Michael S. Valerio[1,9], Christopher L. Dearth[1,7,9] and Luis A. Garza[3,10,11], Major Extremity Trauma Research Consortium (METRC)[8]

In persons with limb loss, prosthetic devices cause skin breakdown, largely because residual limb skin (nonvolar) is not intended to bear weight such as palmoplantar (volar) skin. Before evaluation of treatment efficacy to improve skin resiliency, efforts are needed to establish normative data and assess outcome metric reliability. The purpose of this study was to use optical coherence tomography to (i) characterize volar and nonvolar skin epidermal thickness and (ii) examine the reliability of optical coherence tomography. Four orientations of optical coherence tomography images were collected on 33 volunteers (6 with limb loss) at 2 time points, and the epidermis was traced to quantify thickness by 3 evaluators. Epidermal thickness was greater ($P < .01$) for volar skin (palm) (265.1 ± 50.9 μm, n = 33) than for both nonvolar locations: posterior thigh (89.8 ± 18.1 μm, n = 27) or residual limb (93.4 ± 27.4 μm, n = 6). The inter-rater intraclass correlation coefficient was high for volar skin (0.887−0.956) but low for nonvolar skin (thigh: 0.292−0.391, residual limb: 0.211−0.580). Correlation improved when comparing only 2 evaluators who used the same display technique (palm: 0.827−0.940, thigh: 0.633−0.877, residual limb: 0.213−0.952). Despite poor inter-rater agreement for nonvolar skin, perhaps due to challenges in identifying the dermal−epidermal junction, this study helps to support the utility of optical coherence tomography to distinguish volar from nonvolar skin.

INTRODUCTION

In both the short- and long-terms after amputation, persons with limb loss often experience adverse skin conditions due

[1]Research and Surveillance Section, Extremity Trauma and Amputation Center of Excellence, Defense Health Agency, Falls Church, Virginia, USA; [2]Department of Rehabilitation Medicine, Center for the Intrepid, Brooke Army Medical Center, San Antonio, Texas, USA; [3]Department of Dermatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [4]The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, Maryland, USA; [5]Department of Physical Medicine & Rehabilitation, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA; [6]Department of Physical Therapy, Chapman University, Irvine, California, USA; [7]Department of Rehabilitation, Walter Reed National Military Medical Center, Bethesda, Maryland, USA; [8]Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; [9]Department of Surgery, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA; [10]Department of Cell Biology, Johns Hopkins School of Medicine, Baltimore, USA; and [11]Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, USA

Correspondence: Molly E. Baumann, Research and Surveillance Section, Extremity Trauma and Amputation Center of Excellence, 3551 Roger Brooke Drive, JBSA Fort Sam Houston, Texas 78234, USA. E-mail: Molly.baumann.civ@health.mil

Abbreviations: ICC, intraclass correlation coefficient; OCT, optical coherence tomography

Received 23 October 2023; revised 17 January 2024; accepted 6 February 2024; accepted manuscript published online XXX; corrected proof published online XXX

Cite this article as: JID Innovations 2024;4:100276

to prosthesis use as part of their daily lives and activities. With nearly 2 million people living with limb loss in the United States (Ziegler-Graham et al, 2008), solutions to skin conditions for prosthesis users remain a critical yet unmet need. Among persons with lower limb loss, 73.9% who wear a prosthesis report skin conditions (Koc et al, 2008). Skin conditions range from small abrasions to ulcers, dermatitis, and infections (Meulenbelt et al, 2006). Several factors contribute to skin breakdown, such as moisture and heat surrounding the residual limb along with stresses from the socket (Meulenbelt et al, 2011; Sanders and Daly, 1993). Over time, repeated exposures to this environment, with skin not intended for such conditions, can reduce socket wear time and impede function. This combination can decrease the QOL for persons with limb loss, especially if they are no longer able to wear their prosthesis at all. Even for higher-activity prosthesis users, skin irritation and breakdown throughout the rehabilitation process can impede their return to activity by limiting time in their prosthetic socket. Prior work has shown that lower-extremity prosthesis users doffed their sockets several times over the course of the day likely owing to pain or discomfort or change their liners to limit unfavorable conditions within the socket, even when wearing their prosthesis from 12.8 to 18.8 hours a day (DeGrasse et al, 2023). Finding a solution to prevent skin irritation and breakdown is paramount to continued success of all prosthesis users.

Human skin can exhibit various phenotypes in different body sites, which can enable a difference in stress response and skin condition. The intrinsic properties of and relationship between the epidermis and dermis result in different skin phenotypes (Chang et al, 2002; Garza et al, 2021; Sorrell et al, 2004). Different phenotypes exhibit different properties. Volar skin (palmoplantar) is characterized by greater epidermal thickness, lack of hair follicles, greater keratinocyte cytoplasmic size, and longer collagen length than nonvolar (dorsal) skin (ie, trunk, back) (Figure 1) (Bu et al, 2022; Tsai et al, 2022). These properties make volar skin more resistant to mechanically induced injury and less permeable to irritants and allergens (Farinelli and Berardesca, 2006; Tsai et al, 2022). Thus, volar skin or skin exhibiting these characteristics at the prosthesis interface may enable a different response to prosthesis wear, which could mitigate adverse skin conditions and improve outcomes for prosthesis users.

As nonvolar skin, residual limb skin is not natively adapted to withstand the loads experienced within the prosthetic socket. However, experimental treatment strategies that aim to convert the nonvolar skin of the residual limb to volar skin are being developed in an effort to improve resistance to mechanical wear. Assessing the changes in skin epidermal thickness after treatment will be vital to evaluating treatment efficacy. Because each individual will have different baseline thickness, the change before and after treatment is what is truly of interest (Bailey et al, 2012).

Noninvasive techniques are desirable for quantifying skin thickness among in vitro and in vivo animal models and human studies. Invasive techniques such as biopsies can alter how the skin behaves and include increased risks while leaving a permanent scar. In vitro models requiring a biopsy limit the ability to study the growth of skin unimpeded in

3-dimensional assessment (Sanchez et al, 2022). Utilizing noninvasive techniques in animal models can prevent the killing of an animal to obtain samples for histological analysis (Silver et al, 2012). Noninvasive imaging translates well for human studies to minimize the need for a biopsy and the time required to collect and process the data. Ultrasound devices have been used to assess full-skin thickness and have shown reliability with intraclass correlation coefficients (ICCs) of inter-rater reliability ranging from 0.69 to 0.80 in healthy participants (Peperkamp et al, 2019). However, ultrasound is limited because it assesses the entire skin thickness, including epidermis and dermis, whereas optical coherence tomography (OCT) provides an opportunity to assess the thickness of the epidermis alone.

OCT is a relatively quick and noninvasive imaging modality (Huang et al, 1991; Sattler et al, 2013) likely capable of defining skin thickness with the residual limb and could be utilized as a method of assessing residual limb skin health. This quantification of epidermal thickness has implications across skin science with several clinical applications such as evaluating skin adaptation to wearing prosthetic and orthotic devices. OCT was initially developed for retinal scanning (Huang et al, 1991) but has been further developed and innovatively utilized across many fields, including dermatology (Pierce et al, 2004; Sattler et al, 2013; Zysk et al, 2007). OCT utilizes infrared light, which is sent into the sample. Structures are identified by the time to return to the detector as measured by interferometry (Huang et al, 1991). As the light beam moves across the sample laterally, it creates a 2-dimensional cross-sectional image of the tissue. It is very similar in appearance to a cross-sectional histology slide (Figure 1). Image resolution varies with parameter control of the device itself but typically can visualize 1−2 mm in depth
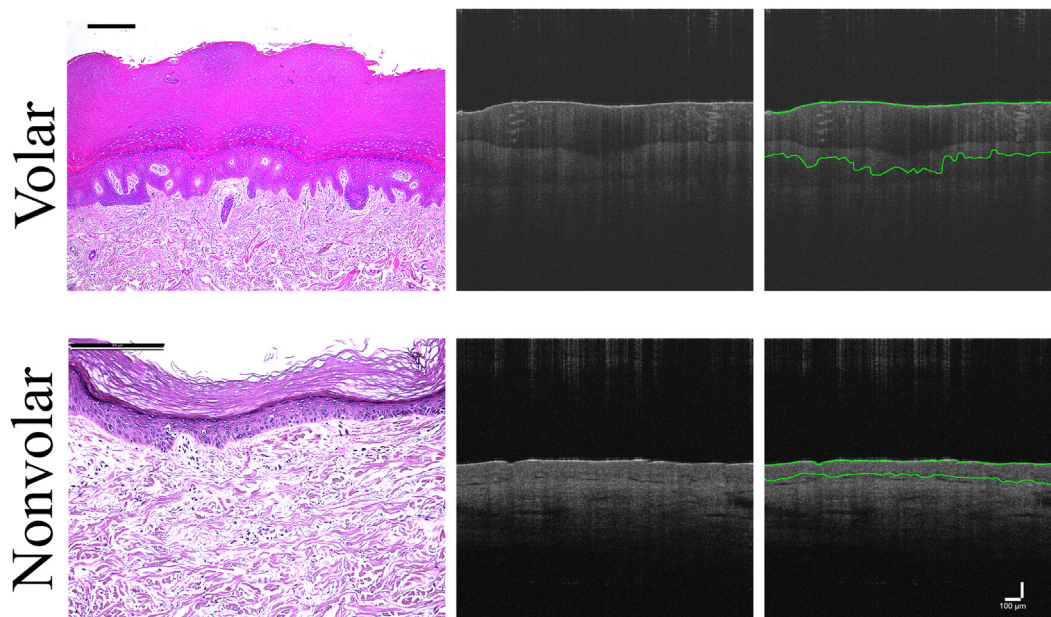

Figure 1. Skin histology and OCT. Comparison of volar with nonvolar skin in histology (left panel) and optical coherence tomography (center and right panel) was performed. Similarities between histology and OCT can be seen, noting a thicker epidermis in volar with both visualization types. Black bar = 200 μm in each histological image. Far right panel demonstrates example tracings of the OCT images. White bar = 100 μm for all OCT images. OCT, optical coherence tomography.

with at least 3 μm spatial resolution (Kollias and Stamatas, 2002; Sattler et al, 2013).

Early use of OCT in human skin qualitatively demonstrated the ability to detect structures and layers of the skin (Welzel et al, 1997). In a murine model, OCT was able to adequately capture the thickness of the epidermis compared with histological samples (Sattler et al, 2013). More recently, OCT has been utilized to visualize residual limb skin of prosthesis users (Swanson et al, 2021). Swanson et al (2021) compared the thickness of residual limb with that of contralateral limb skin and found a greater thickness (although not significant) in the residual limb than in the contralateral limb (n = 3). Additional validation of OCT in measuring residual limb skin thickness is necessary to establish it as a measure for assessing treatment outcomes.

The objective of this study was to evaluate the ability of OCT for quantifying skin thickness in both volar and nonvolar skin. In addition, the objective was to assess inter- and intrarater reliabilities from OCT images collected on the palm (volar skin), posterior thigh (nonvolar skin), and the residual limb of persons with limb loss (nonvolar skin). Given the absence of any intervention, we hypothesized that OCT will differentiate thickness of volar from that of nonvolar skin with good measurement reliability as quantified by the ICC. We hypothesized that there will be no differences in measurements between visit 1 and visit 4 for each measurement location. Results from this work may be used to lay the groundwork for using OCT as a clinical measure within dermatology and skin science to better detect changes in skin for improving outcomes for prosthesis users and beyond.

RESULTS

This study included 33 participants, 6 persons with transtibial limb loss and 27 persons without limb loss (Table 1). OCT

Table 1. Demographic Characteristics of Participants without and with Limb Loss

Characteristic	Total	Persons without Limb Loss	Persons with Limb Loss
Participants, n	33	27	6
Age, y	28.4 ± 8.3	25.6 ± 5.5	41.0 ± 7.4
Gender			
Male	18 (55%)	13 (48%)	5 (83%)
Female	15 (45%)	14 (52%)	1 (17%)
Race/ethnicity			
Non-Hispanic White	17 (58%)	14 (52%)	3 (50%)
Non-Hispanic Black	3 (9%)	2 (7%)	1 (17%)
Hispanic	8 (24%)	6 (22%)	2 (33%)
Other race or >1 race	5 (15%)	5 (19%)	0
Fitzpatrick skin type			
I	4 (12%)	3 (11%)	1 (17%)
II	10 (30%)	10 (37%)	0
III	5 (15%)	3 (11%)	2 (33%)
IV	4 (12%)	3 (11%)	1 (17%)
V	8 (24%)	6 (22%)	2 (33%)
VI	2 (6%)	2 (7%)	0

images were captured according to standardized instructions and settings. Skin thickness was measured on the palm and either posterior thigh or distal end of the residual limb (depending on whether the person had limb loss or not) in 4 orientations at each weekly visit. This yielded 4 images at each skin location for each of 2 visits. Three independent evaluators measured the epidermal thickness. Averages of all thickness measurements in all orientations and among all evaluators were combined. Mean time between measurements was 26.2 ± 7.0 days. Skin at the palm was thicker than skin at the residual limb at visit 1 (261.7 μm vs 95.3 μm, $P < .001$) and visit 4 (268.5 μm vs 91.4 μm, $P < .001$), and skin at the palm was thicker than skin at the posterior thigh at visit 1 (261.7 μm vs 89.4 μm, $P < .001$) and visit 4 (268.5 μm vs 90.2 μm, $P < .001$) (Figure 2). There were no changes in thickness of the thigh or residual limb over time nor between those 2 body sites ($P > .05$); however, palm thickness for orientation 4 images were higher at visit 4 than at visit 1 ($P < .05$). All other palm orientations were not different ($P > .05$) (Figure 2).

ICCs for inter-rater comparisons were used to examine reliability of OCT image assessment stratified by orientation, location, and visit. ICCs were calculated using images collected at both time points and across all 4 orientations. ICC values <0.500 reflect poor reliability, those between 0.500 and 0.750 reflect moderate reliability, those of 0.750–0.900 reflect good reliability, and those >0.900 reflect excellent reliability. The inter-rater ICC for the palm ranged from 0.887 (good) to 0.956 (excellent). The residual limb and thigh had greater variability and lower reliability. The ICC for the residual limb ranged from 0.211 (poor) to 0.580 (moderate), while that of the thigh ranged from 0.292 (poor) to 0.391 (poor) (Figure 3 and Table 2).

Example images of traces with good (Figure 4) and poor (Figure 5) agreement are presented. To evaluate this poor reliability more fully, inter-rater ICCs were calculated for each possible pair of evaluators. Inter-rater ICC values between evaluators 2 and 3 improved (Figure 6 and Table 3). With these 2 evaluators, the palm still exhibited good-to-excellent reliability (range = 0.827–0.940). However, reliability improved for the residual limb and thigh as shown by an increased upper limit of the range of inter-rater ICC values with the residual limb ranging from 0.267 (poor) to 0.952 (excellent), whereas the thigh ranged from 0.633 (moderate) to 0.877 (good) (compare Figure 4 [good agreement] with Figure 5 [poor agreement]). ICC values for evaluators 1 and 2 ranged from 0.855 (good) to 0.952 (excellent) in the palm, −0.019 (poor) to 0.266 (poor) in the residual limb, and 0.055 (poor) to 0.167 (poor) in the thigh. ICC values for evaluators 1 and 3 ranged from 0.972 (excellent) to 0.985 (excellent), 0.006 (poor) to 0.610 (moderate) in the residual limb, and 0.151 (poor) to 0.284 (poor) in the thigh (Table 3).

Intrarater ICC values were calculated over the 4 orientations and stratified by location and visit for each evaluator. ICCs varied on the basis of body location and evaluator. For the palm, all evaluators had an intrarater ICC >0.900 (excellent reliability). For the thigh, all evaluators had an intrarater ICC between 0.750 and 0.900 (good reliability). The ICCs for residual limb ranged from 0.235 for 1 evaluator at visit 4 to 0.921 for 1 evaluator at visit 4, ranging from poor to excellent (Table 4).
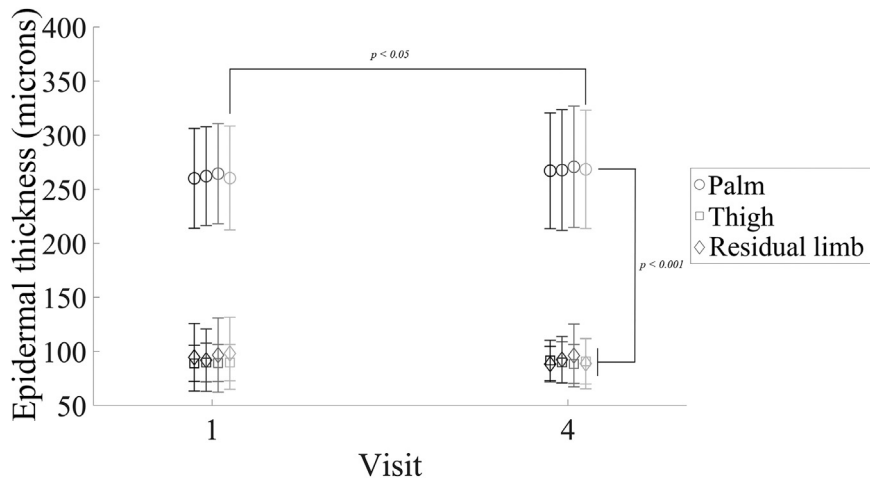
Figure 2. Epidermal thickness of different skin locations. Shown are mean ± SD epidermal thickness for all evaluators at each orientation (paired by grayscale) at visits 1 and 4. Palm epidermal thickness (n = 132 images per time point evaluated by each evaluator) was greater ($P < .001$, 1-way ANOVA with Bonferroni adjustment) than both thigh (n = 108 images per time point evaluated by each evaluator) and residual limb (n = 24 images per time point evaluated by each evaluator) thickness at both time points. There was a change ($P < .05$, paired *t*-test with Bonferroni correction) in epidermal thickness in orientation 4 (lightest gray) of the palm from visit 1 to visit 4. Mean (SD) time between visit 1 and visit 4 was 26.2 ± 7.0 days.

DISCUSSION

This study evaluated the ability of OCT to measure epidermal thickness of volar and nonvolar skin among persons with and without limb loss. Measurements were taken on the thenar eminence of the palm and either posterior thigh or distal end of the residual limb. This study demonstrated that mean thickness varied between volar and nonvolar skin sites when imaged with the OCT and quantified using the methods described below. Reliability of these measurements of epidermal thickness using OCT varied depending on the body site and evaluators involved. Our results demonstrate the importance of image quality and visualization device used for analysis.

Differences in epidermal thickness between volar and nonvolar skin has been found in prior studies (Lintzeri et al, 2022; Mogensen et al, 2008). Mean epidermal thickness of residual limb skin for 3 participants using the OCT was 117 μm (Swanson et al, 2021) compared with 94 μm in this study for residual limb thickness. Epidermal thickness of the calf, which

should relate to the residual limb skin, has been reported to range from 73 to 86 μm (Gambichler et al, 2006; Mogensen et al, 2008). Palm thickness quantified with ultrasound found a mean thickness of 212 μm for females and 221 μm for males (Firooz et al, 2017) compared with 262−269 μm found in this study between visit 1 and visit 4. Although there is a wide range of reported epidermal thicknesses by body region, and comparisons between studies are challenging owing to different population and assessment methods, these prior results demonstrate the capability of this technique and provide results similar to prior work (Lintzeri et al, 2022). Differences in epidermal thickness by anatomical areas, sex/gender, skin phototype, and age are also well-documented (Lintzeri et al, 2022; Sandby-Møller et al, 2003), although others noted no significant differences in epidermal thickness in same anatomical areas across different Fitzpatrick skin types (Mogensen et al, 2008). In this study, mean epidermal thickness values were larger, although not significantly, for the residual limb than the posterior thigh in both visits 1 and 4, similar to the

Figure 3. Inter-rater ICC for 3 evaluators. Shown are mean ± SD for ICC for all 4 orientations for 3 body locations at visit 1 and visit 4. Two-way random effects ANOVA models were used to measure inter-rater reliability, treating participants and evaluators as random effects. Palm (n = 132 images per time point evaluated by each evaluator) demonstrates good-to-excellent reliability, whereas the thigh (n = 108 images per time point evaluated by each evaluator) and residual limb (n = 24 images per time point evaluated by each evaluator) analysis demonstrates poor-to-moderate reliability. ICC, intraclass correlation coefficient.
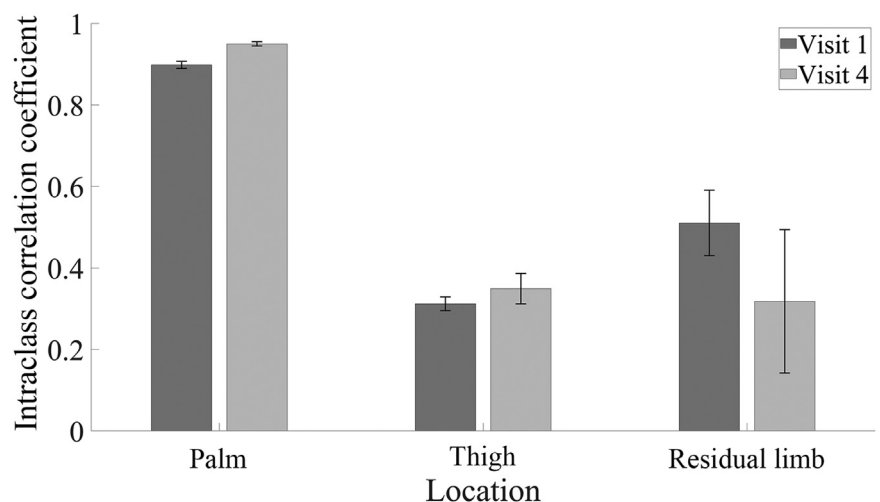
Table 2. Inter-Rater Intraclass Correlation Coefficient for 3 Evaluators

Orientation	Palm Visit 1	Palm Visit 4	Thigh Visit 1	Thigh Visit 4	Residual Limb Visit 1	Residual Limb Visit 4
1	0.903	0.946	0.305	0.315	0.527	0.219
2	0.897	0.956	0.328	0.371	0.393	0.211
3	0.887	0.953	0.324	0.321	0.571	0.58
4	0.907	0.946	0.292	0.391	0.551	0.263

findings from previous work using the OCT in 3 participants comparing the residual with the contralateral limb (Swanson et al, 2021). Our study presented in this paper utilized an indirect comparison owing to measurements of residual limb and posterior thigh being taken on separate individuals, which introduces a new level of variability because skin thickness can vary across individuals.

Several factors contributed to low inter-rater reliability versus intrarater reliability. First, the image analysis process is subjective. Each of the evaluators traced the images manually to determine the epidermal thickness, and it is difficult to identify the exact same demarcation between epidermis and dermis in the image. The deep contrast of the image can make it difficult to identify at times. The volar skin is easier to visualize with its thicker epidermis, perhaps contributing to the increased reliability in volar compared with nonvolar skin sites. When inter-rater reliability ICC values were run for each evaluator pair, the reliability remained high (>0.800 in all cases), whereas the nonvolar skin was more variable, supporting the idea that volar skin is easier to visualize and measure. Second, differences may be attributed to the different visualization device used to trace the images. Image quality and thus the visualization of the dermal−epidermal junction and the epidermal thickness may vary by device. Evaluator 1, who when included demonstrated lower ICC values in the pair comparisons, used an iPad with an apple pencil to manually trace the dermal−epidermal junction, whereas the 2 other evaluators used a laptop and Windows tablet. An iPad can have a smaller screen size than a computer, but it can have a higher pixel density. As a result, the
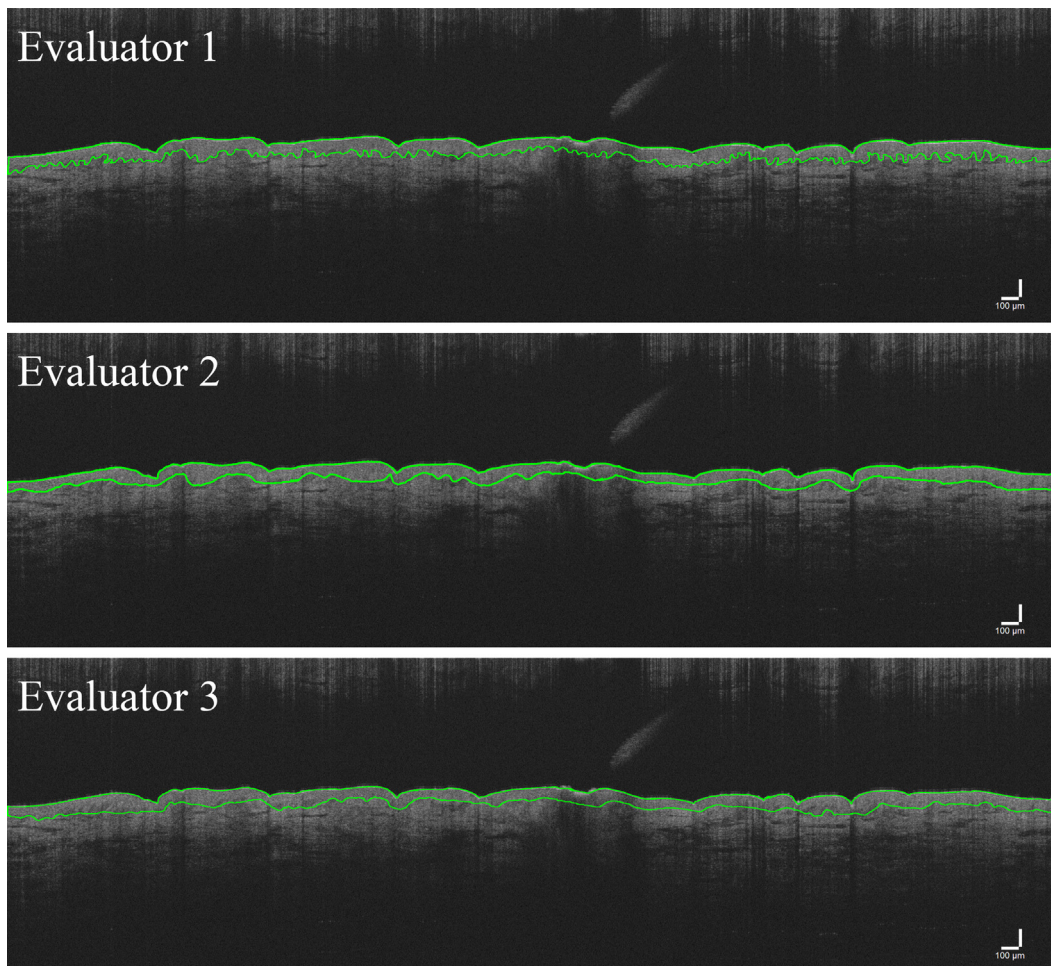

Figure 4. OCT image trace examples of high agreement. Shown are example OCT image traces from all 3 evaluators with good agreement on average thickness of the epidermis. Overall, the 3 traces are very similar in appearance and thickness of quantified measurement. Bar = 100 μm. OCT, optical coherence tomography.
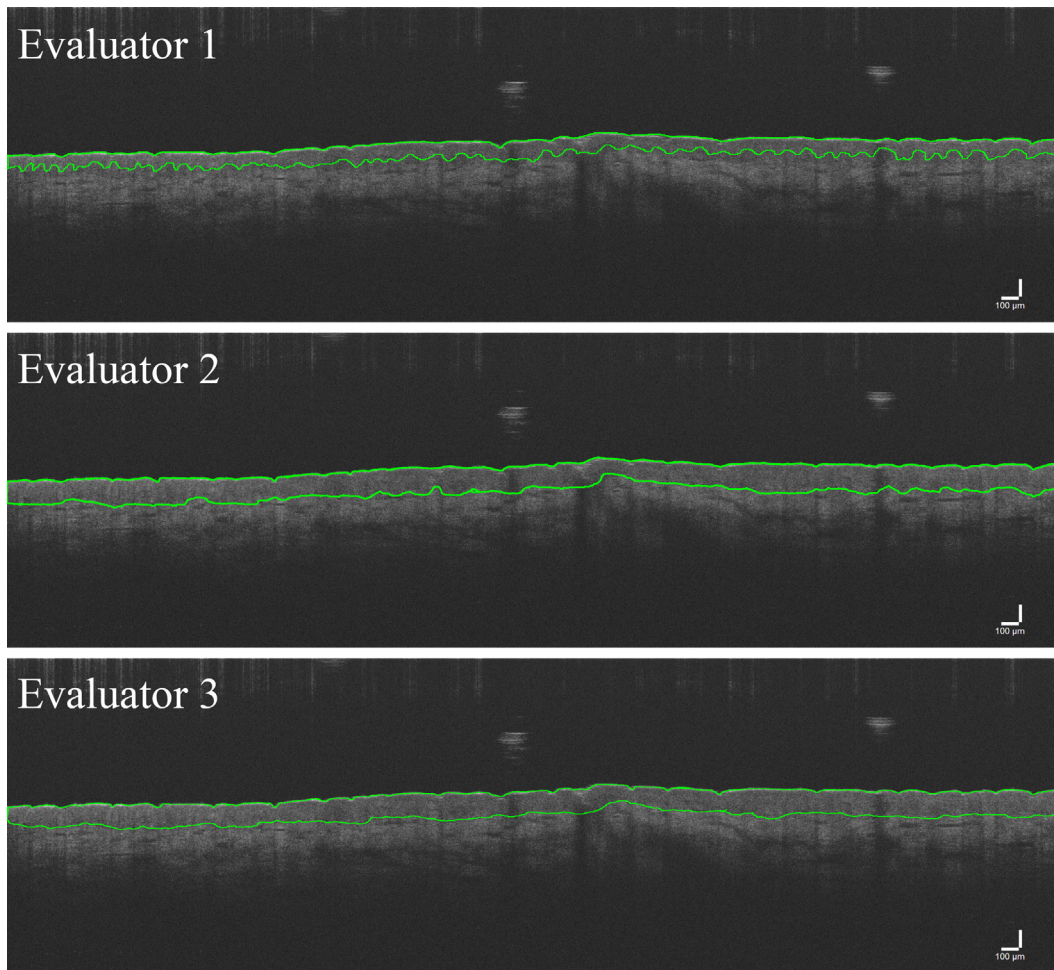
Figure 5. OCT image trace examples of low agreement. Shown are example OCT image traces from all 3 evaluators with poor agreement on average thickness of the epidermis. Top panel demonstrates much thinner tracing of epidermal thickness than middle and lower panel. Bar = 100 μm. OCT, optical coherence tomography.

same image displayed on an iPad or a computer can appear different in size, quality, and clarity. Finally, poor image quality makes it difficult to identify the dermal−epidermal junction, possibly increasing the inter-rater reliability. OCT image quality can be influenced by several factors, including patient cooperation, image acquisition settings, and the presence of artifacts (eg, hair follicle obstructing the view, curvature of area of interest preventing focus on the edges). The ability to maintain the appropriate amount of pressure while capturing the image can influence the overall focus and clarity of the image. Although all image acquisition settings were held constant in this study, the fine focus of the image was adjustable and can result in differing quality of images and thus influence the ability to accurately identify and interpret structures. Improvements in reliability may be achieved by better standardizing the analysis method between the different evaluators to include performing the analysis on the same type of device (whether an iPad or a computer). The visualization device was the main difference between evaluator 1 and the other 2 evaluators. Although there is subjectivity in the analysis of identifying the dermal−epidermal junction, this difference in visualization device can be controlled moving forward and appears to

improve reliability of the assessment. With development of new software-based image analyses, it is probable that an algorithmic approach could be developed in the near future to more consistently and reliably measure epidermal thickness. In addition, implementing an improved training program could help optimize both the quality of the captured images and the analysis process in our future studies. OCT may function as a measure of efficacy in treatment evaluation when acquiring images after thorough training and a detailed SOP with image analysis being performed on the same device by the same evaluator. Standardizing as much of this analysis process as possible can improve agreement between evaluators. In addition, the lack of a true gold standard in noninvasive epidermal thickness quantification makes it difficult to fully assess these ICC values.

There are several limitations to this study. Only 6 of 33 participants had limb loss. With such a small sample size of residual limb images, the variability was greater. This does complicate comparisons between groups. Images of the posterior thigh or calf, which may more closely replicate the residual limb, could be taken on the same participant to be able to compare within an individual for residual limb skin and traditional nonvolar skin. In addition, intrarater reliability
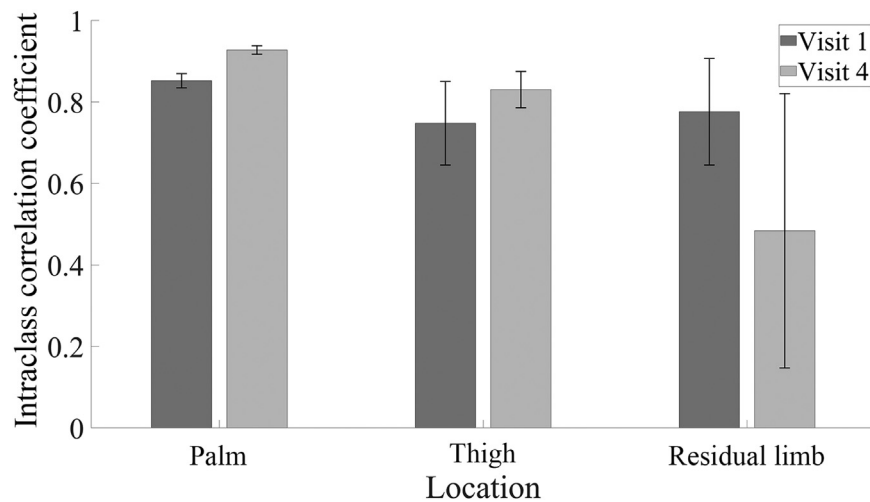
Figure 6. Inter-rater ICC for 2 evaluators. Shown are mean ± SD shown for ICC for all 4 orientations for 3 body locations at visit 1 and visit 4 for only evaluators 2 and 3. Two-way random effects ANOVA models were used to measure inter-rater reliability, treating participants and evaluators as random effects. Compared with ICC values for all 3 evaluators together, palm (n = 132 images per time point evaluated by each evaluator) ICC continues to demonstrate good-to-excellent reliability, whereas thigh (n = 108 images per time point evaluated by each evaluator) and residual limb (n = 24 images per time point evaluated by each evaluator) ICCs have improved from poor to excellent reliability. ICC, intraclass correlation coefficient.

was run on 4 images from the same location in different orientations. Because it is the same area of skin, the expectation is that the variance will not be high but could factor into the intrarater ICC values.

A single orientation of palm images demonstrated an increase in thickness from visit 1 to visit 4. The change (8.1 ± 29.3 µm) is relatively minor compared with the mean thickness of skin at that time point (268.4 ± 54.7 µm) and may not be clinically relevant. Prior studies have shown an ~2 µm change in epidermal thickness over the medioproximal area of the tibia over the course of 2 weeks in 8 participants without limb loss (Swanson et al, 2020). This would be an area of nonvolar skin compared with our area of volar skin. In addition, this 8-µm change occurred over 4 weeks rather than 2. This change seen from visit 1 to visit 4 for a single

orientation is unexpected; however, the change is small in comparison and is similar to changes seen in previous studies. It is possible that the image was not taken in the exact same location and was not perfectly aligned at visit 4.

With the skin being such an adaptive organ, participant occupation could influence their skin thickness. For someone working with their hands for large portions of the day, skin thickness may increase more rapidly than those who do not require their hands to perform manual labor. Job description was not recorded in this study and could have been included as a confounding factor and considered for future studies. Other factors might impact the results in this study. Skin thickness varies by races, ages, and gender. Given the sensitivity of a device such as the OCT, minor differences could be detected by the device. These differences in race,

Table 3. Inter-Rater ICC for Paired Evaluators

Orientation	Palm Visit 1	Palm Visit 4	Thigh Visit 1	Thigh Visit 4	Residual Limb Visit 1	Residual Limb Visit 4
Evaluators 1 and 2 inter-rater ICC values						
1	0.858	0.947	0.124	0.055	0.335	0.035
2	0.857	0.943	0.113	0.154	-0.019	0.243
3	0.855	0.952	0.119	0.105	0.261	0.266
4	0.892	0.947	0.142	0.167	0.12	0.202
Evaluators 1 and 3 inter-rater ICC values						
1	0.983	0.972	0.247	0.189	0.61	0.006
2	0.976	0.985	0.19	0.236	0.203	0.124
3	0.982	0.977	0.238	0.151	0.432	0.145
4	0.976	0.976	0.221	0.284	0.142	0.349
Evaluators 2 and 3 inter-rater ICC values						
1	0.867	0.92	0.707	0.877	0.638	0.504
2	0.859	0.94	0.873	0.841	0.711	0.267
3	0.827	0.933	0.779	0.834	0.818	0.952
4	0.856	0.918	0.633	0.77	0.938	0.213

Abbreviation: ICC, intraclass correlation coefficient.

Table 4. Intrarater ICC Values for All 3 Evaluators

Visit 1	ICC	Visit 4	ICC
Palm			
Evaluator 1	0.952	Evaluator 1	0.966
Evaluator 2	0.960	Evaluator 2	0.976
Evaluator 3	0.960	Evaluator 3	0.958
Thigh			
Evaluator 1	0.835	Evaluator 1	0.832
Evaluator 2	0.885	Evaluator 2	0.896
Evaluator 3	0.876	Evaluator 3	0.898
Residual limb			
Evaluator 1	0.700	Evaluator 1	0.449
Evaluator 2	0.897	Evaluator 2	0.235
Evaluator 3	0.917	Evaluator 3	0.921

Abbreviation: ICC, intraclass correlation coefficient.

age, and gender could have contributed to some of the differences seen and could affect ability to capture reliable images across all populations.

In this study, epidermal thickness of volar skin was different from that of nonvolar (thigh and residual limb) skin. Only a single orientation of images in the palm demonstrated a change between study visits. Overall, the intrarater reliability measures demonstrated the capability of OCT imaging for quantifying epidermal thickness. Although additional efforts can likely improve inter-rater reliability, ultimately, differences between volar and nonvolar skin sites suggest that the current approach may be sufficient to evaluate interventions designed to change nonvolar skin to volar skin. The increased inter-rater ICC values when evaluator 1 was excluded provide insight into the nuances of this method that can impact outcomes such as display devices used for analysis. Future use of this technique to quantify epidermal thickness should consider some modifications to improve the inter-rater ICC values, namely utilizing the same device type to analyze images as well as improved training for image capture and analysis.

MATERIALS AND METHODS

Study participants

Adult volunteers were recruited from 3 military treatment facilities and 1 academic civilian hospital. The study protocol was approved by the institutional review board at each participating site in compliance with all applicable federal regulations governing the protection of human subjects. Participants included those with and without transtibial amputation who were free of skin conditions at the measurement sites. Participants self-reported their gender and race/ethnicity, and researchers assessed Fitzpatrick skin type.

OCT assessment

A handheld Spectral Domain OCT (Ganymede 610C1, Thorlabs) with 930 nm central wavelength, axial scan rate from 5 to 248 kHz, 2.7 mm maximum imaging depth, 6.0 μm axial resolution, 8.0 μm lateral resolution, and sensitivity of 84−102 dB was used for all skin measurements. Skin thickness measurements were taken according to standardized instructions (Supplementary Document A) on 27 participants without an amputation and 6 participants with

transtibial limb loss. Measurements were taken at 2 skin locations approximately once a week for 4 visits to assess measurement reliability over time. For this study, only measurements collected at the initial (visit 1) and final (visit 4) visits were analyzed. For the participants without an amputation, measurements were taken on the thenar eminence of the palm (volar) and posterior thigh (nonvolar). For the participants with a transtibial amputation, measurements were taken at the palm (volar) and residual limb (nonvolar). Care was taken to collect measurements on the same areas of skin during each visit using reference photographs with ruler measurements and skin marker. The skin was marked with a + to ensure proper orientation. Orientation was marked on the OCT software, with a red arrow indicating direction of image capture (Figure 7). Four images were taken to ensure 1 clear image to analyze in 4 orientations at each location site for each visit: north to south (vertical), east to west (horizontal), northwest to southeast (−45), and northeast to southwest (45). These were collected to be able to assess intrarater reliability. Palm measurements were taken over the thenar eminence, thigh images were taken over the posterior mid-thigh, and residual limb images were taken at the most distal end of the limb (0600 location), avoiding the anterior−distal limb above the cut end of the tibia. These 2-dimensional, cross-sectional images were taken with a refractive index of 1.45, length of image of 6 mm, depth of 1.79 mm with a pixel size length at 2.00 μm, and depth at 1.88 μm.

OCT image analysis

Once captured, OCT images were extracted from ThorLabs software and then processed and analyzed on ImageJ (National Institutes of Health). Four images were analyzed in total for each site for each participant, at each visit. Three evaluators with a mixture of skin science and clinical backgrounds independently measured epidermal thickness from OCT images. The contrast and brightness of the extracted images were manually adjusted to optimize visualization of the dermal−epidermal junction. The levels were adjusted specific to each image, but each evaluator analyzed the same adjusted image. This was done to help account for differences seen in image quality across different image takers and body location. This adjustment helped to minimize variation between the evaluators and have the same standardized image analyzed by all 3 evaluators with an optimized dermal−epidermal junction. The dermal−epidermal junction and top of the skin (stratum corneum) were manually traced in ImageJ (Figure 1). To identify the dermal−epidermal junction, evaluators were looking for the line marking the shift from the highly cellular epidermis to the less cellular, collagenous dermis. This is seen by a difference in grayscale intensity within the image. This grayscale difference corresponds with the undulating border between the epidermis and dermis. In addition, blood vessels seen in the images mark the dermal tissue, noting that the junction will be above those blood vessels. The area of the epidermis and image length were both measured. The epidermal thickness was then calculated by dividing the area by the length. In summary, images were collected across 3 body sites, 2 time points, and 4 orientations. Three independent evaluators quantified the thickness for each of these images.

Statistical analysis

All statistical analysis was run utilizing Stata. Comparison between the mean thickness of volar skin (palm) and nonvolar skin (posterior
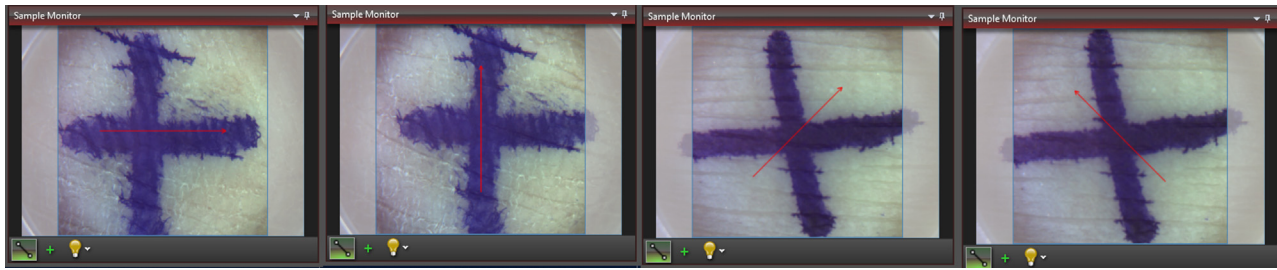
Figure 7. OCT image orientation. Shown are skin marking and image capture software orientation in horizontal, vertical, 45, and −45 orientations. OCT, optical coherence tomography.

thigh or residual limb) was made using 1-way ANOVA with a Bonferroni adjustment for multiple comparisons. Differences in mean skin thickness at visit 1 and visit 4 were evaluated using a paired *t*-test with a Bonferroni correction to account for the 4 orientations measured at each body location. Statistical significance was defined as $P < .05$ for all tests. Two-way random effects ANOVA models were used to measure inter-rater reliability, treating participants and evaluators as random effects. Inter-rater ICC values between the 3 evaluators were calculated by skin location, orientation, and visit. Inter-rater ICCs were stratified by orientation, location, and time. All 3 evaluators quantified each image. ICCs were calculated for each skin location and specific orientation across both visits. ICC values <0.500 were considered poor reliability. Values between 0.500 and 0.750 were considered moderate reliability, those from 0.750 to 0.900 were considered good reliability, and values >0.900 were considered excellent reliability (Koo and Li, 2016; Liljequist et al, 2019). In addition, 2-way random effects ANOVA models were used to measure intrarater reliability, treating participants and orientations as random effects. Intrarater ICC values within the 3 evaluators were calculated over orientation by skin location and visit. Each evaluator quantified the thickness of the 4 orientation images for each skin location and time point. The 4 images were used to assess each evaluator's reliability to themselves. ICC values for each of the 3 evaluators were calculated by skin location and visit.

ETHICS STATEMENT
All procedures were approved by the Johns Hopkins Institutional Review Board, Walter Reed Institutional Review Board, San Antonio Institutional Review Board, and Naval Medical Center San Diego Institutional Review Board. Written, informed consent was obtained prior to initiating any study protocols.

DATA AVAILABILITY STATEMENT
For access to the datasets related to this article, please contact the corresponding author.

ORCIDs
Molly E. Baumann: http://orcid.org/0000-0002-5462-405X
Nina Rossa Haddad: http://orcid.org/0009-0000-1132-6637
Alyssa Salazar: http://orcid.org/0009-0003-9056-8053
W. Lee Childers: http://orcid.org/0000-0002-6119-983X
Shawn Farrokhi: http://orcid.org/0000-0001-9732-7429
Neil B. Goldstein: http://orcid.org/0009-0005-0849-1291
Brad D. Hendershot: http://orcid.org/0000-0001-5400-9551
Lisa M. Reider: http://orcid.org/0000-0002-5373-9197
Richard E. Thompson: http://orcid.org/0000-0001-8378-4426
Michael S. Valerio: http://orcid.org/0000-0003-1261-7078
Christopher L. Dearth: http://orcid.org/0000-0003-3701-0950
Luis A. Garza: http://orcid.org/0000-0002-6547-9695

CONFLICT OF INTEREST
The authors state no conflict of interest.

AUTHOR CONTRIBUTIONS
Conceptualization: SF, BDH, LMR, CLD, LAG; Data Curation: LMR; Formal Analysis: RET; Funding Acquisition: SF, BDH, LMR, CLD, LAG; Investigation: MEB, NRH, AS, NBG, MSV; Methodology: MEB, NRH, AS, WLC, SF, NBG, BDH, LMR, RET, MSV, CLD, LAG; Project Administration: MEB, WLC, SF, NBG, BDH, LMR, MSV, CLD, LAG; Supervision: WLC, SF, BDH, LMR, CLD, LAG; Validation: MEB, NRH, AS, WLC, SF, NBG, BDH, LMR, MSV, CLD, LAG; Visualization: MEB, NRH, AS, RET; Writing - Original Draft Preparation: MEB, NRH, AS; Writing - Review and Editing: MEB, NRH, AS, WLC, SF, NBG, BDH, LMR, RET, MSV, CLD, LAG

DECLARATION OF GENERATIVE ARTIFICIAL INTELLIGENCE (AI) OR LARGE LANGUAGE MODELS (LLMS)
No artificial intelligence or large language model tools were used to prepare this manuscript.

Disclaimer
The views expressed in this manuscript reflect the results of research conducted by the author(s) and do not necessarily reflect the official policy or position of The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., the Defense Health Agency, Department of Defense, nor the United States Government. The identification of specific products or instrumentation is considered an integral part of the scientific endeavor and does not constitute an endorsement or implied endorsement on the part of the authors, Department of Defense, or any component agency.

SUPPLEMENTARY MATERIAL
Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.xjidi.2024.100276.

REFERENCES
Bailey SH, Oni G, Brown SA, Kashefi N, Cheriyan S, Maxted M, et al. The use of non-invasive instruments in characterizing human facial and abdominal skin. Lasers Surg Med 2012;44:131−42.

Bu T, Zhang M, Lee SH, Cheong YE, Park Y, Kim KH, et al. GC-TOF/MS-based metabolomics for comparison of volar and non-volar skin types. Metabolites 2022;12:717.

Chang HY, Chi JT, Dudoit S, Bondre C, van de Rijn M, Botstein D, et al. Diversity, topographic differentiation, and positional memory in human fibroblasts. Proc Natl Acad Sci USA 2002;99:12877−82.

DeGrasse NS, Mertens JC, Brzostowski JT, Allyn KJ, Vamos AC, Krout AJ, et al. Beyond step counts: including wear time in prosthesis use assessment for lower-limb amputation. J Rehabil Assist Technol Eng 2023;10: 20556683231163337.

Farinelli N, Berardesca E. The skin integument: variation relative to sex, age, race, and body region. In: Serup J, Jemec GBE, Grove GL, editors.

Handbook of non-invasive methods and the skin. Abingdon, United Kingdom: Taylor & Francis Group; 2006. p. 27–31.

Firooz A, Rajabi-Estarabadi A, Zartab H, Pazhohi N, Fanian F, Janani L. The influence of gender and age on the thickness and echo-density of skin. Skin Res Technol 2017;23:13–20.

Gambichler T, Matip R, Moussa G, Altmeyer P, Hoffmann K. In vivo data of epidermal thickness evaluated by optical coherence tomography: effects of age, gender, skin type, and anatomic site. J Dermatol Sci 2006;44:145–52.

Garza LA, Lee S, Sweren E, Li A, Kim D, Kim S, et al. 661 cell therapy trial of ectopic fibroblasts to modify skin identity. J Invest Dermatol 2021;141: S115.

Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, et al. Optical coherence tomography. Science 1991;254:1178–81.

Koc E, Tunca M, Akar A, Erbil AH, Demiralp B, Arca E. Skin problems in amputees: a descriptive study. Int J Dermatol 2008;47:463–6.

Kollias N, Stamatas GN. Optical non-invasive approaches to diagnosis of skin diseases. J Investig Dermatol Symp Proc 2002;7:64–75.

Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research [published correction appears in J Chiropr Med 2017;16:346]. J Chiropr Med 2016;15:155–63.

Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation – a discussion and demonstration of basic features. PLoS One 2019;14: e0219854.

Lintzeri DA, Karimian N, Blume-Peytavi U, Kottner J. Epidermal thickness in healthy humans: a systematic review and meta-analysis. J Eur Acad Dermatol Venereol 2022;36:1191–200.

Meulenbelt HE, Dijkstra PU, Jonkman MF, Geertzen JH. Skin problems in lower limb amputees: a systematic review. Disabil Rehabil 2006;28: 603–8.

Meulenbelt HE, Geertzen JH, Jonkman MF, Dijkstra PU. Skin problems of the stump in lower limb amputees: 1. A clinical study. Acta Derm Venereol 2011;91:173–7.

Mogensen M, Morsy HA, Thrane L, Jemec GB. Morphology and epidermal thickness of normal skin imaged by optical coherence tomography. Dermatology 2008;217:14–20.

Peperkamp K, Verhulst AC, Tielemans HJP, Winters H, van Dalen D, Ulrich DJO. The inter-rater and test-retest reliability of skin thickness and skin elasticity measurements by the DermaLab Combo in healthy participants. Skin Res Technol 2019;25:787–92.

Pierce MC, Strasswimmer J, Park BH, Cense B, De Boer JF. Advances in optical coherence tomography imaging for dermatology. J Invest Dermatol 2004;123:458–63.

Sanchez MM, Orneles DN, Park BH, Morgan JT. Automated epidermal thickness quantification of in vitro human skin equivalents using optical coherence tomography. BioTechniques 2022;72:194–200.

Sandby-Møller J, Poulsen T, Wulf HC. Epidermal thickness at different body sites: relationship to age, gender, pigmentation, blood content, skin type and smoking habits. Acta Derm Venereol 2003;83:410–3.

Sanders JE, Daly CH. Normal and shear stresses on a residual limb in a prosthetic socket during ambulation: comparison of finite element results with experimental measurements. J Rehabil Res Dev 1993;30:191–204.

Sattler E, Kästle R, Welzel J. Optical coherence tomography in dermatology. J Biomed Opt 2013;18:061224.

Silver R, Helms A, Fu W, Wang H, Diaconu D, Loyd CM, et al. Using optical coherence tomography for the longitudinal non-invasive evaluation of epidermal thickness in a murine model of chronic skin inflammation. Skin Res Technol 2012;18:225–31.

Sorrell JM, Baber MA, Caplan AI. Site-matched papillary and reticular human dermal fibroblasts differ in their release of specific growth factors/cytokines and in their interaction with keratinocytes. J Cell Physiol 2004;200: 134–45.

Swanson EC, Friedly JL, Wang RK, Sanders JE. Optical coherence tomography for the investigation of skin adaptation to mechanical stress. Skin Res Technol 2020;26:627–38.

Swanson EC, Friedly JL, Wang RK, Sanders JE. Optical coherence tomography for the investigation of skin adaptation in lower limb prosthesis users. J Prosthet Orthot 2021;33:255–65.

Tsai J, Rostom M, Garza LA. Understanding and harnessing epithelial–mesenchymal interactions in the development of palmoplantar identity. J Invest Dermatol 2022;142:282–4.

Welzel J, Lankenau E, Birngruber R, Engelhardt R. Optical coherence tomography of the human skin. J Am Acad Dermatol 1997;37:958–63.

Ziegler-Graham K, MacKenzie EJ, Ephraim PL, Travison TG, Brookmeyer R. Estimating the prevalence of limb loss in the United States: 2005 to 2050. Arch Phys Med Rehabil 2008;89:422–9.

Zysk AM, Nguyen FT, Oldenburg AL, Marks DL, Boppart SA. Optical coherence tomography: a review of clinical development from bench to bedside. J Biomed Opt 2007;12:051403.