Check for updates

Received: 1 August 2021 | Accepted: 25 September 2021

DOI: 10.1111/srt.13113

ORIGINAL ARTICLE

WILEY

Skin measurement devices to assess skin quality: A systematic review on reliability and validity

Mirte Langeveld[1] | **Lara S. van de Lande**[2] | **Eimear O' Sullivan**[2] | **Berend van der Lei**[3] | **Joris A. van Dongen**[4]

[1] Medical Faculty, Erasmus University Medical Center, Rotterdam, the Netherlands

[2] Craniofacial Unit Great Ormond Street Hospital for Children and UCL Great Ormond Street Institute of Child Health, London, UK

[3] Department of Plastic Surgery, University of Groningen and University Medical Center Groningen, Groningen, the Netherlands

[4] Department of Plastic Surgery, University Medical Center Utrecht, Utrecht, the Netherlands

Correspondence
Joris A. van Dongen, MD, Department of Plastic Surgery, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands.
Email: jorisavandongen@gmail.com

Abstract

Background: Many treatments aim to slow down or reverse the visible signs of skin aging and thereby improve skin quality. Measurement devices are frequently employed to measure the effects of these treatments to improve skin quality, for example, skin elasticity, color, and texture. However, it remains unknown which of these devices is most reliable and valid.

Materials and methods: MEDLINE, Embase, Cochrane Central, Web of Science, and Google Scholar databases were searched. Instruments were scored on reporting construct validity by means of convergent validity, interobserver, intraobserver, and interinstrument reliability.

Results: For the evaluation of skin color, 11 studies were included describing 16 measurement devices, analyzing 3172 subjects. The most reliable device for skin color assessment is the Minolta Chromameter CR-300 due to good interobserver, intraobserver, and interinstrument reliability. For skin elasticity, seven studies assessed nine types of devices analyzing 290 subjects in total. No intra and interobserver reliability was reported. Skin texture was assessed in two studies evaluating 72 subjects using three different types of measurement devices. The PRIMOS device reported excellent intra and interobserver reliability. None of the included reviewed devices could be determined to be valid based on construct validity.

Conclusion: The most reliable devices to evaluate skin color and texture in ordinary skin were, respectively, the Minolta Chromameter and PRIMOS. No reliable device is available to measure skin elasticity in ordinary skin and none of the included devices could be determined to be designated as valid.

KEYWORDS
facial surgery, measurement tools, plastic surgery, skin quality, skin rejuvenation, systematic review

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Skin Research and Technology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

The skin is the largest and most visible organ of the human body with important protective and regulatory functions. The epidermal barrier forms the first line of defense against exogenic factors and pathogenic microorganisms. The skin also plays an important role in thermoregulation, metabolic processes, and sensory perception.[1–3] Adjacent to its function, skin plays a key role in aesthetics; unfortunately, skin quality decreases over time due to aging, especially in the face. Facial aging is characterized by many changes in a broad spectrum of facial skin features, for example, pigmentation, wrinkles, and rosacea.[4,5] Aging of skin can be categorized into two types of aging: intrinsic and extrinsic aging. Intrinsic aging derives from genetic and hormonal influences, whereas extrinsic aging is caused by environmental factors, such as cigarette smoke, ultraviolet radiation, or trauma.[4,6] In the epidermis of the skin, aging of the face is characterized by loss of dermal mast cells and fibroblasts as well as by shortening of telomeres. In the dermis, lower levels of collagen, dysfunctional collagen, and a reduction of elastin fibers are observed.[4] These cellular changes result in increased pigmentation, loss of elasticity, and formation of wrinkles over time.[5,7]

Nowadays, people have become progressively concerned about their aged facial skin features. Many autologous treatments, for example, lipofilling, platelet-rich plasma, or nanofat, aim to either slow down or reverse these visible signs of skin aging and thereby improving skin quality.[8] Generally, skin quality and skin quality improvement is assessed merely by visual inspection by the patient and practitioner, which is accompanied by disadvantages of interperson variability and recall-bias, making the results rather unreliable. Some clinicians determine the effectiveness of such interventions by assessing skin quality with the use of a measurement tool as, for example, tristimulus colorimetry to measure skin color, the Cutometer or Ballistometer for skin elasticity, and polarization imaging techniques to assess skin texture.[9–11] However, it remains unknown whether these devices are accurate and dependable. Therefore, the aim of this study is to systematically search for the best-validated medical devices to assess skin quality (i.e., skin color, texture, and elasticity) in the most reliable way.

2 | METHODS

2.1 | Protocol, information sources, and search

This systematic review was performed according to the PRISMA statement.[12] The databases MEDLINE, Embase, Cochrane Central, Web of Science, and Google Scholar were searched on April 16, 2019. An update search was performed on December 15, 2020. The detailed search strategy is provided in the Supplementary Content (S1).

2.2 | Eligibility criteria and study selection

Title and abstract were independently screened by two authors (M.L. and L.v.d.L.) using eligibility criteria. Full article studies were included if studies investigated the reliability and validity of medical devices assessing changes in human "ordinary" aged skin, that is, skin color, texture, or elasticity (Table 1). Studies were included if reported at least one of the following items regarding skin quality measurement devices: intraobserver reliability, interobserver reliability, interinstrument reliability, or construct validity. Studies evaluating content and criterion validity were not found. Studies assessing the quality of "diseased" skin, for example, melanoma, scars, or burn wounds, were excluded as well as animal studies. Reference lists of included studies were hand-searched for relevant studies. Disagreements were discussed during a consensus meeting with the last author (J.v.D.).

2.3 | Assessment of quality of included studies and risk of bias

The included studies were graded on quality of evidence using the Oxford Center for Evidence-Based Medicine (OCEBM) criteria.[13] Disclosure agreements and funding status were reviewed for each study.

2.4 | Data extraction

Measurement devices were scored on reporting construct validity by means of convergent validity and inter or intraobserver as well as interinstrument reliability. For construct validity, the Pearson's correlation coefficients of correlations between measurement devices were extracted and the median was depicted in a correlogram. Correlations > 0.5 or < -0.5 were considered strong. For reliability, intraclass correlation coefficients (ICCs) were reported. ICCs > 0.8 were considered good, moderate between 0.6 and 0.8, and poor < 0.6.

3 | RESULTS

3.1 | Included studies

The initial search identified 3724 publications (Figure 1). The update search yielded 621 additional publications. Hand-searching reference lists of included publications identified two additional records. After abstract screening, 4296 were excluded. Fifty studies were read in full text and assessed on eligibility criteria. Twenty-seven studies did not describe an outcome of interest and were excluded. Four publications were reviews and therefore excluded. One publication was excluded because of evaluating diseased skin. One study was excluded as it was a letter to the editor. Following full-text assessment, 18 publications were included in this systematic review.[9,10,14–29]

3.2 | Study characteristics

3.2.1 | Skin color

Eleven studies assessed skin color describing a total of 16 different measurement devices analyzing 3172 subjects (Table 2).[9,14–22,25,30] The largest study by Uter et al. accounted for 2287 of included

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Human skin	Diseases and trauma affecting skin quality, for example, burn wounds, scars, and disease-caused
Medical devices assessing human skin texture, color, or elasticity	
Reporting of intraobserver and/or interobserver reliability and/or interinstrument observer reliability and/or validity	
Prospective and retrospective studies	Case reports, conference abstracts, letter to the editor, and reviews

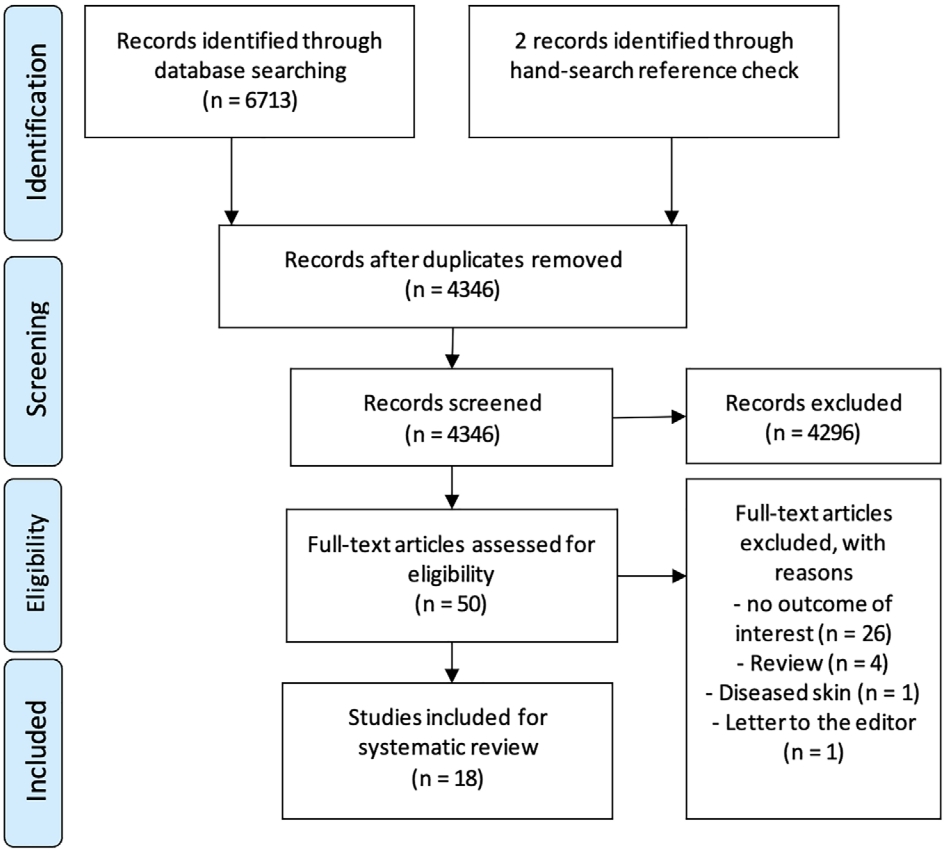

FIGURE 1 Flow diagram of study selection

subjects.[9] All studies evaluated measurement devices in a predominantly Caucasian population, except for one study by Wright et al.[14] Wright et al. researched the DRS probe and Mexameter MX 18 in a predominantly (68.5%) African American population ($n = 503$).[14]

The two most frequently employed techniques were narrow-band reflectance spectrophotometry and tristimulus colorimetry. In narrow-band reflectance spectrophotometry, differences in red and near infrared light absorption and reflection of hemoglobin and melanin are used to measure vascularization (erythema) and pigmentation (melanin) of the skin.[28,31] Included devices using reflectance spectrophotometry to assess skin color were the Mexameter MX 16 and 18, DermaLab Combo, DSM II ColorMeter, and DermaSpectrometer.[14–19,22] In Tristimulus colorimetry, white LED is scattered in all directions and the reflected light is measured by the probe. The reflected light is analyzed and expressed in the L*a*b* color system and Individual Typology Angle index values (ITA). L* expresses brightness on the black-white axis, a* expresses erythema values on the red-green axis, and b* gives the color position on the blue-yellow axis.[15] Instruments using Tristimulus colorimetry to measure skin color are the Minolta Chromameter CR-200 and CR-300, Colorimeter CL-400, PhotoVolt ColorWalk Colorimeter, and Visi-Chroma VC-100.[9,15,19–21]

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Study characteristics of studies on skin color measurement

Author, year	Population (*n*)	Device	Principle	Clinical parameter	Measurement region	Intervention	Measurement timings	Repetitive measurements (*n*)
Wright et al., 2016	503 African American	DRS Probe Mexameter MX 18	Diffuse reflectance spectroscopy Narrow-band reflectance spectrophotometry	Melanin, erythema Melanin, erythema	Inner part of upper arm Inner part of upper arm	–	Baseline Baseline	3 3
Matias et al., 2015	30	Antera 3D Mexameter MX 18 Colorimeter CL-400	Reflectance mapping with L*a*b* color system Narrow-band reflectance spectrophotometry Tristimulus colorimetry with L*a*b* color system	Melanin, erythema, skin color Melanin, erythema Skin color	The back The back The back	UVB light exposure at various intensities UVB light exposure at various intensities UVB light exposure at various intensities	Baseline, 2, 7, 12, and 14 days Baseline, 2, 7, 12, and 14 days Baseline, 2, 7, 12, and 14 days	5 5 5
Baquie and Kasraee, 2014	12	Dermacatch Mexameter MX 16	Visible-spectrum reflectance colorimeter Narrow-band reflectance spectrophotometry	Melanin, erythema Melanin, erythema	Volar side of the forearm and the back Volar side of the forearm and the back	UVB light exposure, methyl nicotine cream or dermocorticoid cream UVB light exposure, methyl nicotine cream or dermocorticoid cream	Baseline, 2, 7, and 14 days Baseline, 2, 7, and 14 days	10 10

(Continues)

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Author, year	Population (*n*)	Device	Principle	Clinical parameter	Measurement region	Intervention	Measurement timings	Repetitive measurements (*n*)
Hua et al., 2014	20	"Soft Plus" with melanin probe Mexameter MX 18	Double wavelength reflectance photometry Narrow-band spectrophotometry	Melanin Melanin	Face Face	–	Baseline Baseline	3-5 3–5
Gankande et al., 2014	30	DermaLab Combo	Narrow-band reflectance spectrophotometry	Melanin, erythema	Head, neck, chest, back, arm, leg	–	Baseline	3
Uter et al., 2013	2287	Minolta Chromameter CR-300 Reflektometer RM 100	Tristimulus colorimetry with Yxy color system Remission photometry	Skin color Skin reflectance	Inner part of upper arm Inner part of upper arm	–	Baseline Baseline	3 3
Van der Wal et al., 2013	50	Mexameter MX 18 Colorimeter CL-400 DSM II ColorMeter	Narrow-band reflectance spectrophotometry Tristimulus colorimetry with L*a*b* color system Narrow-band reflectance spectrophotometry and tristimulus colorimetry with L*a*b* color system	Melanin, erythema Skin color Skin color, erythema melanin	Trunk, upper and lower extremities Trunk, upper and lower extremities Trunk, upper and lower extremities	–	Baseline Baseline Baseline	2 2 2

(Continues)

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Author, year	Population (*n*)	Device	Principle	Clinical parameter	Measurement region	Intervention	Measurement timings	Repetitive measurements (*n*)
Bailey et al., 2012	88	Chromometer	Principle not mentioned	Pigmentation	Forehead, midcheek, jawline, neck, and abdomen	–	Baseline	–
Barel et al., 2001	12	Visi-Chroma VC-100 Minolta Chromameter CR-200	Tristimulus colorimetry with L*a*b* color system Tristimulus colorimetry with L*a*b* color system	Skin color Skin color	– –	DHA 5% cream, methyl nicotine cream or sodium lauryl sulfate exposure DHA 5% cream, methyl nicotine cream or sodium lauryl sulfate exposure	Baseline, 2, 4, and 24 h Baseline, 2, 4, and 24 h	10 10
Kerckhove et al., 2001	60	Minolta Chromameter CR-300	Tristimulus colorimetry with L*a*b* color system	Skin color	Ventral side of the forearm	–	Baseline, 7 days	–
Shriver et al., 2000	80	PhotoVolt Color-Walk colorimeter DermaSpectrometer	Tristimulus colorimetry Narrow-band reflectance spectrophotometry	Skin color with L*a*b* color system Melanin, erythema	Inner part of the upper arm, forehead Inner part of the upper arm, forehead	–	Baseline Baseline	3 3

L*a*b* = Commision International d'Eclairage (CIE) color system. Colors are represented by three variables: L*, the lightness-darkness axis; a* the red-green axis; and b*, the blue-yellow axis.
Yxy = Commision International d'Eclairage (CIE) color system. Y value represents lightness-darkness axis.
DHA = dihydroxyacetone, product used for tanning of the skin.

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2.2 | Skin elasticity

For skin elasticity, seven studies assessed nine types of measurement devices analyzing 290 subjects in total (Table 3).[10,17,23–26] The Cutometer SEM 575 and Cutometer MPA 580 were the most frequently used devices and were assessed in four studies.[10,17,24,26] The Cutometer MPA 580 is currently still available for purchase, while the Cutometer SEM 575 has been discontinued. The Cutometer uses a suction and optical measuring system to measure various parameters, such as skin distensibility (R0), gross elasticity (R2), and skin firmness (R7).[26] Xu et al. assessed the 3D-DIC, which measures the displacement of skin and minor as well as major strain of skin deformation using unidirectional force.[24] Other measurement devices include the BTC-2000, which measures elastic deformation of skin under subatmospheric pressure, and the Ballistometer BLS780, which uses an impact and indentation measuring system.[10,25] Peperkamp et al. evaluated the Dermalab Combo to measure skin elasticity through suction.[23] Hua et al. assessed the Soft Plus with an elasticity probe, which also measures skin elasticity by measuring stress under suction application.[17] Lastly, in a single study, elastography with the Toshiba iAplio 900 was used to measure skin elasticity by measuring the velocity of ultrasonic waves through skin tissue.[29]

3.2.3 | Skin texture

Skin texture was assessed in two studies evaluating 72 subjects using three different types of measurement devices: the Visioscan VC 98, the PRIMOS, and the PRIMOS^{lite} (Table 4).[11,27,28] The PRIMOS and PRIMOS^{lite} devices use rapid in vivo evaluation of the skin (PRIMOS) to measure surface roughness. This technique is based on the deflection of projected parallel stripe patterns on the skin due to differences in skin surface profile. The Visioscan VC 98 is a UVA-light camera that measures roughness with the Surface Evaluation for Living Skin method (SELS). The PRIMOS^{lite} is a portable version of the PRIMOS.

3.3 | Reliability

3.3.1 | Skin color

Interobserver reliability was highest for the Minolta Chromameter CR-300 (Table 5). Van den Kerckhove et al. reported intraclass coefficients between 0.92 and 0.99 in 60 patients with measurements provided by two independent observers.[21] Both Van den Kerckhove et al. and Uter et al. reported good intraobserver reliability for the Minolta Chromameter as well (ICC 0.98–0.99 and 0.926–0.954, respectively).[9,21] Intraobserver reliability for the Reflektometer RM 100 was good in a large cohort of 2287 subjects (ICC 0.938–0.946).[9] In a single study of 50 participants, Van der Wal et al. assessed the interobserver reliability of the Mexameter MX 18, Colorimeter CL-400, and DSM II ColorMeter.[19] The Mexameter MX 18 and DSM II ColorMeter achieved good interobserver reliability (ICC 0.92–0.94 and 0.89–0.96, respectively). The Colorimeter CL-400 achieved moderate to good interobserver reliability (ICC 0.79–0.97).[32] Gankande et al. reported interobserver reliability of the DermaLab Combo assessing both melanin and erythema. ICCs for erythema were poor to moderate (ICC 0.54–0.73) and good for melanin (ICC 0.91–0.95).[18] Intraobserver reliability was not tested for the Mexameter MX 18, DSM II ColorMeter, ColoriMeter CL-400, and DermaLab Combo.

3.3.2 | Skin elasticity

Intraobserver reliability was tested for the Toshiba iAplio ultrasonography (Table 5).[29] Good intraclass coefficients were reported between 0.842 and 0.987 for three repeated measurements (Table 5). Interobserver reliability was not tested. The DermaLab Combo was the only device that reported interobserver reliability, intraclass coefficients were poor to moderate and varied between 0.23 and 0.76 with measurements repeated by two different observers.[23]

3.3.3 | Skin texture

The PRIMOS reported interobserver reliability of 0.85–0.88, with measurements by three observers in 60 patients (Table 5).[28] Intraobserver reliability of the PRIMOS was 0.96–0.99.[28] The Visioscan VC 98 achieved interobserver reliability of 0.95–1.00 in 12 subjects, with measurements repeated by three different observers.[27] In the same study, interobserver reliability coefficients of the PRIMOS^{lite} ranged between 0.35 and 1.00 (Table 5).[27] Intraobserver reliability was not reported for the Visioscan VC 98 and PRIMOS^{lite}.

3.4 | Validity

3.4.1 | Skin color

Darkness and erythema measurements of devices were correlated with the Fitzpatrick skin-type scale for four devices (Figure 2). Correlation for darkness values and Fitzpatrick skin-type scale score was significant for the Chromometer (R = 0.78), Colorimeter CL-400 (R = –0.68), DSM II Colormeter (R = 0.7), and Mexameter MX 18 (R = 0.72).[19,25] Three devices were assessed for correlation between erythema values and Fitzpatrick skin-type scale score. No significant correlations were found for the Colorimeter CL-400 (R = 0.12), DSM II ColorMeter (0.44), and Mexameter MX 18 (R = 0.44).[19]

Construct validity was tested most frequently for the Mexameter MX 18 (Figures 2 and 3). Measurements of skin darkness by the Mexameter MX 18 were significantly correlated with darkness values of the Antera 3D (R = 0.73), DRS Probe (R = 0.88), and "Soft plus" with melanin probe (R = 0.96, Figure 2).[19] Measurements for skin erythema were significantly correlated with measurements of the Antera 3D (R = 0.77) (Figure 3). The previous model of the Mexameter, the Mexameter MX 16, reported significant correlations for both melanin and erythema values with the Chromameter (R = = –0.77, R = 0.76), the

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Study characteristics of studies on skin elasticity measurement

Author, year	Population (*n*)	Device	Principle	Clinical parameter	Measurement region	Measurement timings	Repetitive measurements (*n*)
Peperkamp et al., 2019	49	DermaLab Combo	Vertical suction	ViscoElasticity (VE), Young's elasticity modulus (E), and skin retraction time (R) ViscoElasticity (VE), Young's elasticity modulus (E), and skin retraction time (R) Viscoelasticity (VE), Young's elasticity modulus (E), skin retraction time (R)	Six locations on arm	Baseline, 45 min	2
Xu et al., 2019	12	3D-DIC Cutometer MPA 580	Deformation of skin under unidirectional force Suction and optical measuring system	Displacement of skin, minor strain, major strain Net elasticity (R5), skin firmness (R7), total recovery (R8)	Volar forearm Volar forearm	Baseline Baseline	3 3
Paluch et al., 2020	57	Toshiba iAplio 900 Ultrasonograph	Shear wave elastography	Tissue strain measured by velocity of ultrasonic wave propagation	Face	Baseline	3
Hua et al., 2014	20	"Soft Plus" with elasticity probe Cutometer MPA 580	Stress/deformation of skin by suction application Suction and optical measuring system	Elasticity Skin distensibility (R0)	Face Face	Baseline Baseline	3 3
Woo et al., 2014	20	Cutometer MPA 580 Ballistometer BLS780	Suction and optical measuring system Impact and indentation measuring system	Skin distensibility (R0), return to original skin (R1), gross elasticity (R2), last maximal amplitude (R3), last minimal amplitude (R4), net elasticity (R5), viscoelasticity (R6), skin firmness (R7), total recovery (R8) Firmness and elasticity	Forehead, cheek, and volar forearm Forehead, cheek, and volar forearm	Baseline Baseline	3 3
Bailey et al., 2012	88	BTC-2000	Deformation of skin under subatmospheric pressure	Elastic deformation and stiffness	Forehead, midcheek, jowl, neck, and abdomen	Baseline	–
Ahn et al., 2007	44	Cutometer SEM 575 Moiré topography image	Suction and optical measuring system Visual evaluation of digital contour lines (scale 1–5)	Skin distensibility (R0), gross elasticity (R2), net elasticity (R5), viscoelasticity (R6), skin firmness (R7), total recovery (R8) Contour lines	Cheek Cheek	Baseline Baseline	– –

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Study characteristics of studies on skin texture measurement

Author, year	Population (*n*)	Device	Principle	Clinical parameter	Measurement region	Intervention	Measurement timings	Repetitive measurements (*n*)
Kottner et al., 2012	12	Visioscan VC 98 PRIMOS^{lite}	Phaseshift rapid evaluation Phaseshift rapid evaluation	Surface roughness Surface roughness	Volar forearmVolar forearm	–	Baseline Baseline	33
Bloemen et al., 2011	60	PRIMOS	Phaseshift rapid evaluation	Surface roughness	Trunk, arm, leg, or head	–	Baseline	2

TABLE 5 Reliability of assessed devices

		Reliability		
Author, year	Device	Intraobserver (ICC) range	Interobserver (ICC) range	Interinstrument (ICC) range
Color				
Kerckhove et al., 2001	Minolta Chromameter CR-300	0.98–0.99	0.92–0.99	0.99–0.999
Uter et al., 2013	Minolta Chromameter CR-300 Reflektometer RM 100	0.926–0.954[a]0.938–0.946	–	–
Van der Wal et al., 2013	Mexameter MX 18 Colorimeter CL-400 DSM II ColorMeter	–	0.92–0.94 0.79–0.97 0.89–0.96	–
Gankande et al., 2014	DermaLab Combo	–	0.54–0.95	–
Elasticity				
Paluch et al., 2020	Toshiba iAplio 900 Ultrasonograph	0.842–0.987	–	–
Peperkamp et al., 2019	DermaLab Combo	–	0.23–0.76	–
Texture				
Kottner et al., 2012	Visioscan VC 98 PRIMOS^{lite}	–	0.95–1.00 0.35–1.00	–
Bloemen et al., 2011	PRIMOS	0.96–0.99	0.85–0.88	–

[a]The reliability of the Minolta Chromameter CR-300 was tested in smaller cohorts of 190, 10, and 8 patients.

Dermacatch ($R = 1$, $R = 1$), and the DermaSpectrometer ($R = 0.53$, $R = 0.81$).

The Minolta Chromameter CR-200 showed significant correlation with darkness values of the Visi-Chroma VC-100 ($R = 0.93$). The Minolta Chromameter CR-300 had significant correlation with the Reflektometer RM 100 ($R = 0.69$). However, both the Minolta Chromameter CR-200 and CR-300 have currently been discontinued, while the newer model Chromameter CR-400 has not yet been evaluated in clinical research.

3.4.2 | Skin elasticity

Ahn et al. reported significant correlation between Cutometer values and the values on a digital grading scale, the so-called Moiré topography ($R = 0.67$).[26] Moiré topography is a digital program which generates contour lines on a digital photograph of a patient. An evaluator clinically rates these contour lines from 1 to 5 for decreasing skin elasticity.[26] Moreover, the Cutometer reported significant correlations with measurements of the Soft plus with elasticity probe and the 3-DIC ($R = -0.64$ and 0.57, respectively, Figure 4).[24] No significant correlation was found between the Cutometer MPA 580 and the Ballistometer BLS780 for gross elasticity (R2), net elasticity (R5), and skin firmness (R7) parameters ($R < 0.5$).[10]

3.4.3 | Skin texture

The measurements of the Visioscan VC 98 and PRIMOS^{lite} were not significantly correlated, so no construct validity could be determined.[27]

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

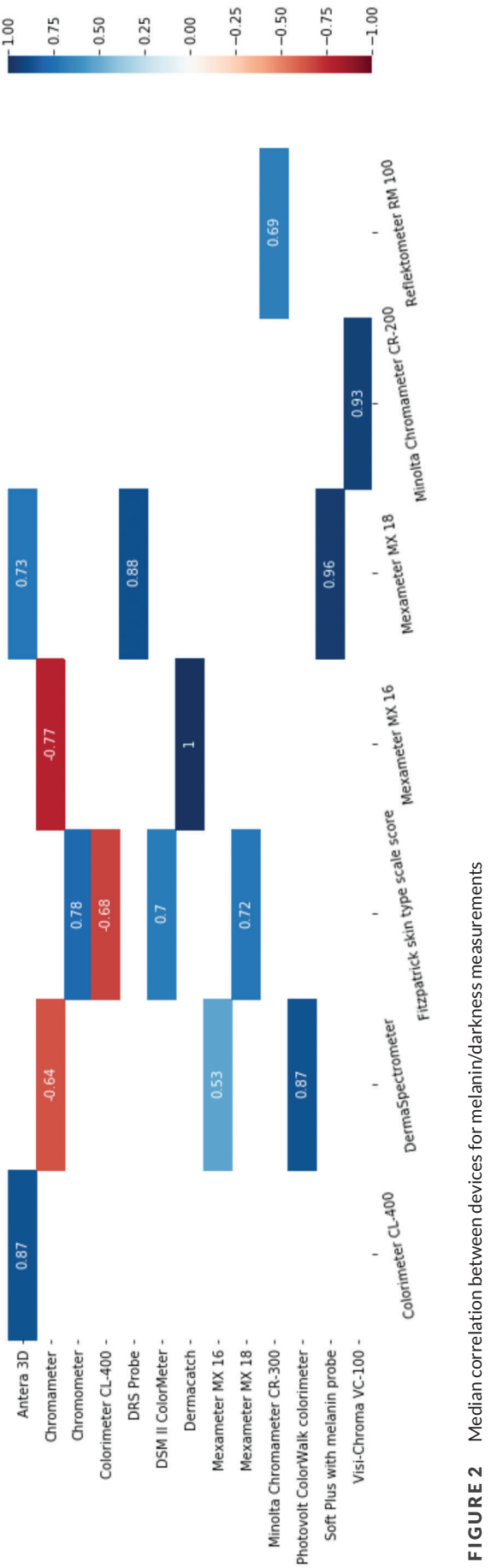

FIGURE 2 Median correlation between devices for melanin/darkness measurements

3.5 | Conflict of interest/risk of bias

All 18 studies were level of evidence III studies.[9,10,14–29] Eight out of 18 articles did not report funding and conflict of interest.[15–17,21,24,26,27,29,30] However, one of these studies was performed by a contract research organization, which developed the investigated device.[15] Three articles reported no conflict of interest, but did not elaborate on funding.[11,19,25] Seven articles disclosed funding.[9,10,14,18,20,22,23] In six of these studies, no conflict of interest was apparent. In a single study, three authors were involved with the company that developed the investigated measurement device.[20]

4 | DISCUSSION

This review aimed to determine the most reliable and validated available medical devices for assessing skin color, texture, and elasticity. The most reliable medical device for skin color evaluation is the Minolta Chromameter CR-300 due to good interobserver, intraobserver, and interinstrument reliability in a Caucasian population. The most reliable medical device for skin texture evaluation is the PRIMOS device with excellent intra and interobserver reliability. For the evaluation of skin elasticity, no device could be designated as superior because none of the included devices reported at least both intra and interobserver reliability. Unfortunately, none of the reviewed measurement devices for skin color, elasticity, or texture could be designated to be superior in terms of validity based on construct validity. Yet, many other critical aspects of validity that are needed to determine which measurement device is best valid, were missing in all included studies. These aspects are content and criteria validity, responsiveness, and interpretability of the included devices. Hence, none of the devices to measure skin color, texture, or elasticity could be selected as valid.

Validity can be divided into construct validity, content validity, and criterion validity. Construct validity by means of convergent validity is the degree to which different devices that should theoretically measure the same construct are actually correlated. The convergent validity of the included measurement devices of skin color and elasticity demonstrated that the majority of these devices measure similar constructs, for example, the Mexameter MX 18 measurements showing significant correlation with the Antera 3D, DRS Probe, and "Soft plus" with melanin probe. When multiple devices interlink, a web of correlations or correlogram can be constructed. Theoretically, this increases the probability that what is being measured is valid. A limitation of the use of convergent validity is that a correlation between devices does not automatically mean that either of the devices actually measures the intended parameter. For example, in case of skin color, van der Wal et al. correlate measurements of the Mexameter MX 18, Colorimeter CL-400, and DSM II ColorMeter to the Fitzpatrick skin phototyping scale. The Fitzpatrick scale is widely used to categorize skin color but was originally developed to select the correct dose of photochemotherapy in the treatment of psoriasis.[33] The parameters assessed in the Fitzpatrick scale are, therefore, primarily focused on predicting the reaction of the skin to ultraviolet light. This scale might, therefore, not

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

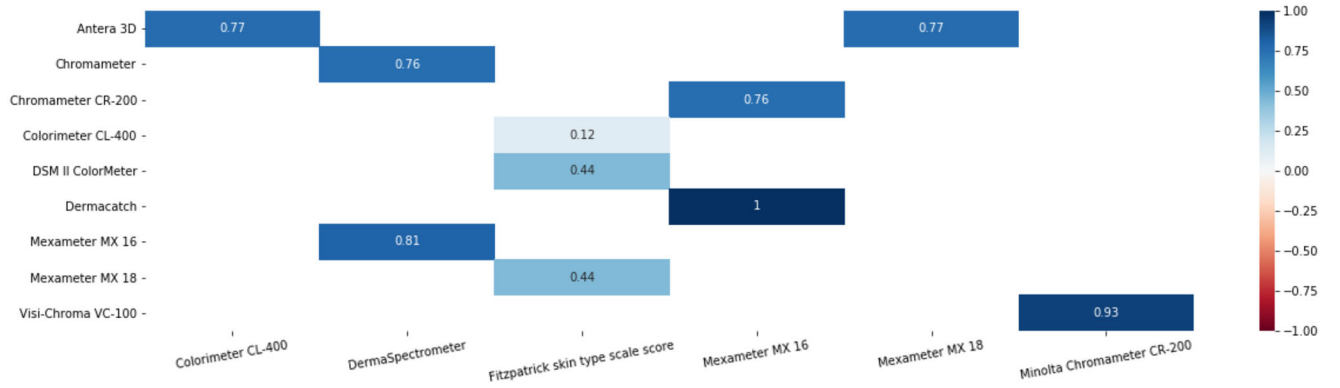

FIGURE 3 Median correlation between devices for erythema measurements

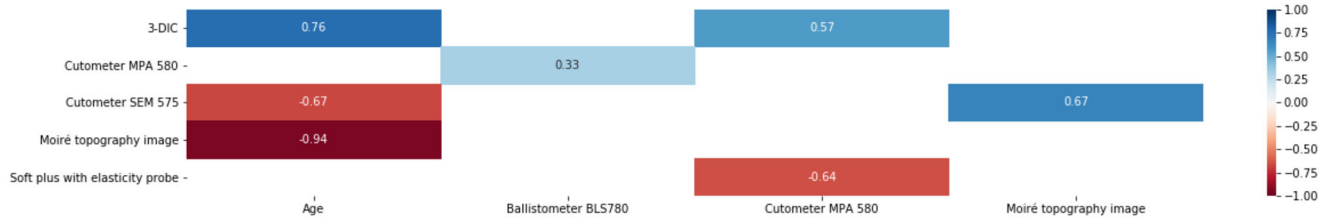

FIGURE 4 Median correlation between devices for elasticity measurements

reflect all facets of the parameter "color" like undertones, redness, or evenness in color. Therefore, it is not surprising that Van der Wal et al. did not find a significant correlation between erythema values of the aforementioned included measurement devices and Fitzpatrick skin-type scale score. This highlights the shortcomings of the use of only construct validity.

Content validity specifies the degree to which assessment instruments are representative of and relevant to the targeted construct these devices are designed to measure. For content validity, a measurement device should measure all aspects of a construct, for example, skin color. By default, a measurement device cannot factor all different nuances of skin color. Criterion validity refers to the degree to which a measure relates to an outcome. Generally, this concerns comparing the instrument under assessment to a different instrument that has been considered as valid, that is, the "gold standard." To date, there is no gold standard for the evaluation of skin color, elasticity, and texture, making criterion validity assessment impossible.

Besides different validity criteria, a measurement device is only useful in clinical practice when it detects clinically meaningful changes in the measured parameter (responsiveness). Responsiveness can be evaluated by correlating changes in the values of measurement devices following intervention to clinical subjective scores. These scores could be questionnaires or scales documenting the perceived benefit of the intervention from the patients' and clinicians' viewpoint.[34] The evaluation of responsiveness is critical to determine whether a device can detect clinically meaningful changes in skin quality. Moreover, a clinical understanding of what the quantitative values or changes in value of the device mean (interpretability) should also be investigated. None of these parameters were reported in any of the included studies.

Our analysis focuses on devices that evaluate the outcome of interventions that influence skin quality. Therefore, in the context of evaluation of outcomes of clinical interventions like lipofilling, we propose the following criteria for a measurement device to be considered valid. First, a device should be reliable, meaning both inter and intraobserver reliability should be tested and ICC should be at least 0.8. Second, a measurement device should be able to detect a clinically relevant change of a measurable parameter. Clinical relevance can be detected by correlating changes in measurement values to clinical results assessed by practitioners or subjects following cosmetic intervention like lipofilling. These patient- or practitioner-reported outcome measures could, for instance, be blinded clinical photographic analyses, or satisfaction as measured with the FACE-Q questionnaires.[35] Reproducibility, responsiveness, and interpretability are critical aspects for a device to be considered valid and reliable.

5 | CONCLUSION

The most reliable devices to evaluate skin color and texture in ordinary aged skin were, respectively, the Minolta Chromameter and PRIMOS. No reliable device is available to measure skin elasticity in ordinary aged skin and none of the included devices could be determined to be designated as valid. Independent responsiveness and interpretability research of available devices is needed to determine which

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

device measures skin quality in the most reliable and reproducible way.

ACKNOWLEDGMENTS

The authors received no specific funding for this work. None of the authors has a financial interest in any of the products, or devices mentioned in this manuscript.

CONFLICT OF INTEREST

The authors have no conflict of interest to disclose.

REFERENCES

1. Boer M, Duchnik E, Maleszka R, Marchlewicz M. Structural and biophysical characteristics of human skin in maintaining proper epidermal barrier function. Postepy Dermatol Alergol. 2016;33(1):1–5. https://doi.org/10.5114/pdia.2015.48037
2. Rosso JD, Zeichner J, Alexis A, Cohen D, Berson D. Understanding the epidermal barrier in healthy and compromised skin: clinically relevant information for the dermatology practitioner: Proceedings of an Expert Panel Roundtable Meeting. J Clin Aesthet Dermatol. 2016;9(4 Suppl 1):S2–8.
3. Yaar M, Gilchrest BA. Skin aging: postulated mechanisms and consequent changes in structure and function. Clin Geriatr Med. 2001;17(4):617-30. https://doi.org/10.1016/s0749-0690(05)70089-6
4. Kosmadaki MG, Gilchrest BA. The role of telomeres in skin aging/photoaging. Micron. 2004;35(3):155-9. https://doi.org/10.1016/j.micron.2003.11.002
5. Fisher GJ, Wang ZQ, Datta SC, Varani J, Kang S, Voorhees JJ. Pathophysiology of premature skin aging induced by ultraviolet light. N Engl J Med. 1997;337(20):1419-29. https://doi.org/10.1056/NEJM199711133372003
6. Freedberg IM, Eisen, AZ, Wolff K, Austen KF, Goldsmith LA, Katz SI, et al. Fitzpatrick's dermatology in general medicine. McGraw-Hill; 2003.
7. Kang S, Fisher GJ, Voorhees JJ. Photoaging: pathogenesis, prevention, and treatment. Clin Geriatr Med. 2001;17(4):643-59. https://doi.org/10.1016/s0749-0690(05)70091-4
8. van Dongen JA, Langeveld M, van de Lande LS, Harmsen MC, Stevens HP, van der Lei B. The effects of facial lipografting on skin quality: a systematic review. Plast Reconstr Surg. 2019;144(5):784e-97e. https://doi.org/10.1097/PRS.0000000000006147
9. Uter W, Benz M, Mayr A, Gefeller O, Pfahlberg A. Assessing skin pigmentation in epidemiological studies: the reliability of measurements under different conditions. Skin Res Technol. 2013;19(2):100-6. https://doi.org/10.1111/srt.12013
10. Woo MS, Moon KJ, Jung HY, Park SR, Moon TK, Kim SN, et al. Comparison of skin elasticity test results from the Ballistometer((R)) and Cutometer((R)). Skin Res Technol. 2014;20(4):422-8. https://doi.org/10.1111/srt.12134
11. Bargo PR, Kollias N. Measurement of skin texture through polarization imaging. Br J Dermatol. 2010;162(4):724-31. https://doi.org/10.1111/j.1365-2133.2010.09639.x
12. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264-9. https://doi.org/10.7326/0003-4819-151-4-200908180-00135
13. OCEBM Levels of Evidence Working Group. The Oxford Levels of Evidence 2. Accessed August 8th, 2020, https://www.cebm.net/2016/05/ocebm-levels-of-evidence/
14. Wright CY, Karsten AE, Wilkes M, Singh A, du Plessis J, Albers PN, et al. Diffuse reflectance spectroscopy versus Mexameter((R)) MX18 measurements of melanin and erythema in an African population. Photochem Photobiol. 2016;92(4):632-6. https://doi.org/10.1111/php.12607
15. Matias AR, Ferreira M, Costa P, Neto P. Skin colour, skin redness and melanin biometric measurements: comparison study between Antera((R)) 3D, Mexameter((R)) and Colorimeter((R)). Skin Res Technol. 2015;21(3):346-62. https://doi.org/10.1111/srt.12199
16. Baquie M, Kasraee B. Discrimination between cutaneous pigmentation and erythema: comparison of the skin colorimeters Dermacatch and Mexameter. Skin Res Technol. 2014;20(2):218-27. https://doi.org/10.1111/srt.12109
17. Hua W, Xie H, Chen T, Li L. Comparison of two series of non-invasive instruments used for the skin physiological properties measurements: the 'Soft Plus' from Callegari S.p.A vs. the series of detectors from Courage & Khazaka. Skin Res Technol. 2014;20(1):74-80. https://doi.org/10.1111/srt.12086
18. Gankande TU, Duke JM, Danielsen PL, DeJong HM, Wood FM, Wallace HJ. Reliability of scar assessments performed with an integrated skin testing device — the DermaLab Combo((R)). Burns. 2014;40(8):1521-9. https://doi.org/10.1016/j.burns.2014.01.025
19. van der Wal M, Bloemen M, Verhaegen P, Tuinebreijer W, de Vet H, van Zuijlen P, et al. Objective color measurements: clinimetric performance of three devices on normal skin and scar tissue. J Burn Care Res. 2013;34(3):e187-94. https://doi.org/10.1097/BCR.0b013e318264bf7d
20. Barel AO, Clarys P, Alewaeters K, Duez C, Hubinon JL, Mommaerts M. The Visi-Chroma VC-100: a new imaging colorimeter for dermatocosmetic research. Skin Res Technol. 2001;7(1):24-31. https://doi.org/10.1034/j.1600-0846.2001.007001024.x
21. Van den Kerckhove E, Staes F, Flour M, Stappaerts K, Boeckx W. Reproducibility of repeated measurements on healthy skin with Minolta Chromameter CR-300. Skin Res Technol. 2001;7(1):56-9. https://doi.org/10.1034/j.1600-0846.2001.007001056.x
22. Shriver MD, Parra EJ. Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry. Am J Phys Anthropol. 2000;112(1):17-27. https://doi.org/10.1002/(SICI)1096-8644(200005)112:1<17::AID-AJPA3>3.0.CO;2-D
23. 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(6):787-92. https://doi.org/10.1111/srt.12718
24. Xu Z, Dela Cruz J, Fthenakis C, Saliou C. A novel method to measure skin mechanical properties with three-dimensional digital image correlation. Skin Res Technol. 2019;25(1):60-7. https://doi.org/10.1111/srt.12596
25. 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(2):131-42. https://doi.org/10.1002/lsm.21147
26. Ahn S, Kim S, Lee H, Moon S, Chang I. Correlation between a Cutometer and quantitative evaluation using Moire topography in age-related skin elasticity. Skin Res Technol. 2007;13(3):280-4. https://doi.org/10.1111/j.1600-0846.2007.00224.x
27. Kottner J, Schario M, Garcia Bartels N, Pantchechnikova E, Hillmann K, Blume-Peytavi U. Comparison of two in vivo measurements for skin surface topography. Skin Res Technol. 2013;19(2):84-90. https://doi.org/10.1111/srt.12009
28. Bloemen MC, van Gerven MS, van der Wal MB, Verhaegen PD, Middelkoop E. An objective device for measuring surface roughness of skin and scars. J Am Acad Dermatol. 2011;64(4):706-15. https://doi.org/10.1016/j.jaad.2010.03.006
29. Paluch L, Pietruski P, Noszczyk B, Kwiek B, Ambroziak M. Intra-rater reproducibility of shear wave elastography in the evaluation of facial

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

skin. Postepy Dermatol Alergol. 2020;37(3):371-6. https://doi.org/10.5114/ada.2018.81144
30. Masuda Y, Yamashita T, Hirao T, Takahashi M. An innovative method to measure skin pigmentation. Skin Res Technol. 2009;15(2):224-9. https://doi.org/10.1111/j.1600-0846.2009.00359.x
31. Diffey BL, Oliver RJ, Farr PM. A portable instrument for quantifying erythema induced by ultraviolet radiation. Br J Dermatol. 1984;111(6):663-72. https://doi.org/10.1111/j.1365-2133.1984.tb14149.x
32. Adler JE, Nico L, Van de Vord P, Skoff AM. Modulation of neuropathic pain by a glial-derived factor. Pain Med. 2009;10(7):1229-36.
33. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol. 1988;124(6):869-71. https://doi.org/10.1001/archderm.124.6.869
34. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care. 2001;17(4):479-87.
35. Klassen AF, Cano SJ, Schwitzer JA, Baker SB, Carruthers A, Carruthers J, et al. Development and psychometric validation of the FACE-Q skin, lips, and facial rhytids appearance scales and adverse effects checklists for cosmetic procedures. JAMA Dermatol. 2016;152(4):443-51. https://doi.org/10.1001/jamadermatol.2016.0018

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Langeveld M, van de Lande LS, O' Sullivan E, van der Lei B, van Dongen JA. Skin measurement devices to assess skin quality: A systematic review on reliability and validity. Skin Res Technol. 2022;28:212–224. https://doi.org/10.1111/srt.13113

16000846, 2022, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/srt.13113 by Erasmus University Rotterdam Universiteitsbibliotheek, Wiley Online Library on [30/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License