Published in final edited form as:

Skin Res Technol. 2013 November ; 19(4): 484–491. doi:10.1111/srt.12072.

Intra- and inter-rater reliability of digital image analysis for skin color measurement

Marilyn Sommers¹, Barbara Beacham¹, Rachel Baker², and Jamison Fargo³

¹Nursing, University of Pennsylvania, Philadelphia, PA, USA

²Patient Services, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

³Psychology, Utah State University, Logan, UT, USA

Abstract

Background—We determined the intra- and inter-rater reliability of data from digital image color analysis between an expert and novice analyst.

Methods—Following training, the expert and novice independently analyzed 210 randomly ordered images. Both analysts used Adobe® Photoshop lasso or color sampler tools based on the type of image file. After color correction with Pictocolor® in camera software, they recorded L*a*b* (L*=light/dark; a*=red/green; b*=yellow/blue) color values for all skin sites. We computed intra-rater and inter-rater agreement within anatomical region, color value (L*, a*, b*), and technique (lasso, color sampler) using a series of one-way intra-class correlation coefficients (ICCs).

Results—Results of ICCs for intra-rater agreement showed high levels of internal consistency reliability within each rater for the lasso technique (ICC = 0.99) and somewhat lower, yet acceptable, level of agreement for the color sampler technique (ICC = 0.91 for expert, ICC = 0.81 for novice). Skin L*, skin b*, and labia L* values reached the highest level of agreement (ICC 0.92) and skin a*, labia b*, and vaginal wall b* were the lowest (ICC = 0.64).

Conclusion—Data from novice analysts can achieve high levels of agreement with data from expert analysts with training and the use of a detailed, standard protocol.

Keywords

skin imaging; skin color; digital image analysis; inter-rater reliability; image analysis; protocols

Analysis of skin color is useful to understand skin phenotype (1), compare injured and non-injured skin (2–4), describe pathology (5), and monitor the outcomes from therapeutic interventions (6). Changes in skin color over time are monitored in dermato-cosmetic research; the color response may indicate the integrity of the skin barrier, skin sensitivity, properties of drugs and other vehicles applied to the skin, and skin protection from sun screens (7).

Capture of digital images provides the opportunity for skin and tissue color analysis when more advanced techniques such as spectrophotometry (7–9) are not available or feasible. Digital image analysis has been used for over 15 years in medical imaging (10, 11) and is a

well-established scientific technique to measure skin color (12–14) and skin injury (15–17). To ensure the accuracy and precision of data, collected through digital image color analysis (DICA), strategies are needed to assess if training and the use of a standard protocol lead to data that are repeatable for the same analyst and reproducible across different analysts. Reliability testing also ensures that the fidelity of analysis procedures is maintained over the lifetime of a study.

In the context of a study of genital skin injury after consensual sexual intercourse as compared with rape, we developed and tested a protocol with a series of 210 digital images of constitutive and facultative skin color at five anatomical sites as well as mucous membrane and vaginal wall color. The purpose of this study was to determine the intra- and inter-rater reliability of data from an experienced color analyst (the expert) and a novice color analyst who was trained using a structured protocol for DICA of the skin.

Background

The foundation of color measurement (colorimetry) occurred in 1931, when the International Commission on Illumination (CIE, or Commission Internationale de l'Eclairage) system was established (18–20). Color scientists recognized the need for uniform mathematical models to describe how colors might be represented. These models, or color spaces, are one-, two-, or three-dimensional geometric models of the colors that can be seen. A color space represents color in terms of intensity values and visually is often represented by three-dimensional shapes such as cubes, cones, or polyhedra (18, 21).

The need for a uniform color space for color measurement led to a number of transformations that resulted in the CIE 1976 ($L^*a^*b^*$) color space, or CIELAB (CIE $L^*a^*b^*$). The CIELAB represents colors relative to a white reference point. The advantage of the CIELAB color space over other models is that perceived differences between any two colors is proportional to the geometric distance in the color space between these two color values (18, 20, 22). In the CIELAB color space, values L^* , a^* , and b^* are plotted at right angles to one another, to form a three-dimensional coordinate system. Equal distances in the space represent equal color differences (21). Value L^* represents light/dark and extends from 0 (black) to 100 (white). Value a^* represents the red/green axis; positive a^* is red and negative a^* is green. Value b^* represents the yellow/blue axis; positive b^* is yellow and negative b^* is blue. While theoretically there are no specific numerical limits for a^* and b^* (18), in practice a^* and b^* have upper and lower limits depending on the analysis technique used (23, 24).

Human skin color results from unpigmented skin color mixed with colors of various constitutive pigments such as melanins, hemoglobins, and carotenes. The major dominant biological pigments are red coloration from hemoglobin in the microcapillary vascular network of the skin and melanin, which provides varying degrees of brown coloration at the skin surface (23). Skin color is found in the positive a^* and positive b^* quadrants of the CIELAB color space (25, 26).

Methods

We used pre-existing digital images with no subject identifiers from a database obtained from the principal investigator's ongoing work determining the role of skin color in detecting skin injury following consensual sexual intercourse and rape. The prospective work was approved by the affiliated university's institutional review board and all subjects signed consent.

Magnified skin images were captured with a Leica® DFC420 C camera (Leica Microsystems, Wetzlar, Germany) interfaced with a Cooper Surgical (Cooper Surgical, Trumbull, CT, USA) Leisegang® (Leisegang, Berlin, Germany) colposcope system. The DFC420 C is a high resolution, Digital FireWire Camera with an advanced 5 megapixel sensor that digitizes information of the charge-coupled device (CCD) in the camera head, leading to optimum noise suppression. The resulting images are highly detailed and appropriate for life sciences applications. For this study, the camera captured uncompressed, digital images in the full tagged image file format (tiff) that were six megapixels or larger for analysis and post-capture processing. Images were captured in a light- and temperature-controlled laboratory after a 10 min rest period to allow subjects to acclimate to the environment.

The database contained digital images from more than 300 participants who underwent a baseline sexual assault forensic examination prior to engaging in consensual sexual intercourse. Each image folder contained the following images from seven sites: 1) volar forearm, one inch distal from the antecubital space; 2) inner upper arm, two inches distal from the axilla; 3) inner upper thigh two inches below the groin (inguinal area); 4) sacrum; 5) external genitalia; 6) inner vaginal wall; and 7) one untanned, unexposed site (constitutive skin color) on the thigh, back, or sacrum as determined by the camera operator. Additionally, each folder contained an image of a Munsell® (ColorAccuracy.com, Lincroft, NJ, USA) Color Checker mini card for use during color correction (see below).

The expert has completed more than 1000 image analyzes over 12 years and was trained by a doctorally prepared color scientist from a skin cosmetics company and by scientists at a university color institute. Prior to intra- and inter-rater reliability testing, the expert and novice completed five training sessions. The first two sessions focused on practice sessions with a written protocol (4), and the second two sessions focused on review of images and guidance from the expert on the location for color analysis. At the final session the novice and expert compared data from practice sessions, reviewed images, and discussed strategies to revise and refine the protocol for consistency. Following protocol revision, the novice used the protocol to practice DICA with approximately 300 sample images.

Expert and novice reliability testing

For final reliability testing, subject folders were selected on the basis of their completeness and clarity of the images. Ten folders were selected by the expert with seven images each. To enable testing for intra-rater (within rater) reliability, a master file was created with each of the ten folders repeated three times (7 images per folder × 30 folders), thereby creating 210 images for DICA. To control for order effect, images in the folders were placed in the random order. DICA of all images in the folder was completed by the expert rater. Then DICA of the randomly ordered 210 images was performed by the novice without discussion or collaboration with the expert. All analyzes were performed on Levono® (Levono, Beijing, China) ThinkPad laptop computers.

Digital image color analysis procedures

To perform the color analysis, we used Adobe PhotoShop® Creative Suite 4 (CS4) Premium. Using Photoshop's® color sampler tools, one can retrieve and record $L^*a^*b^*$ color values. Color modes are user-selectable and include the CIE $L^*a^*b^*$ color space measurement system. Pictocolor® in Camera v.4.5 (27) is a software package that works as a plug-in within Adobe Photoshop® CS4 Premium. In the case of Picto-color® in Camera, this plug-in works within Adobe Photoshop® to adjust the color profile and provide accurate color to the industry/reference standard. This correction is done with the use of an industry-standard color card such as the Munsell® Color Checker mini card (28). The Munsell®

Color Checker is an object of reference with ‘true colors’ necessary for color correction prior to sampling for $L^*a^*b^*$ values. While ‘true color’ is the theoretical goal, measurement error precludes ever obtaining ‘true color’. The Munsell® Color Color Checker is a 3.25×2.25 card with a scientifically prepared checkerboard array of 24 colored squares with known color values. Many of the squares represent natural objects of interest, such as human skin and blue sky (28). The mini card was photographed prior to each set of participant images, illuminated by the same lighting as the participant, and located near a skin site of interest. Thus, each participant folder had a color card image, which was used to create an individual color profile to be applied to all seven images in the folder for the purpose of color correction.

Color correction is a process by which an image of the reference standard such as the Color Checker card allows the image analyst to make color and white balance adjustments that bring the image closer to the ‘true color’ (See Fig. 1). Images are color-corrected to ‘true color’ standards using the Pictocolor® in Camera software. The color sampling performed with Adobe Photoshop® is a common method used by researchers for color correction and digital image color standardization (29, 30).

Color sampling techniques

Adobe PhotoShop® CS4 provides two sampling strategies for DICA: the lasso and the eye-dropper color sampler tools. The ‘lasso tool’ allows the analyst to select an area for color analysis with a range of pixels appropriate to the goals of the DICA. For consistency we analyzed an area of 120,000–150,000 pixels to provide a representative sample of skin color (4) for images of the forearm, upper arm, thigh, sacrum, and untanned areas. To increase our consistency, we viewed the image as a clock and assessed the sampling area clockwise beginning at 12 o’clock. When we reached an area of representative color for the entire sample, we used the lasso tool to encircle an area of 120,000–150,000 pixels and recorded the mean L^* , a^* , and b^* values from the expanded histogram window. We repeated this process two additional times for each image for a total of three L^* , a^* , and b^* color values for each image. We chose to sample a total of three times based on our previous work on skin color analysis (4, 5).

Images with smaller areas of representative skin color, the genital and vaginal views in particular, do not have large areas of uninterrupted skin available for DICA because of anatomy. For these areas, we used the Adobe Photoshop® eye-dropper color sampler tool set at the ‘ 5×5 Average’ pixel sample size (total of 25 pixels) for consistency. We again used the strategy of sampling beginning at 12 o’clock on the image and, moving clockwise until we found a representative area, recorded the mean L^* , a^* , and b^* values from the expanded histogram window. We repeated the color sampler measurements eight more times for a total of nine color measurements. In our previous work we have found that sampling a small area nine times provides representative color analysis (4, 5).

Data transformation

Color data from the Adobe Photoshop® color sampler technique can be used without transformation. When using the Adobe Photoshop® lasso technique, the 0–100 range of L^* values has been scaled by Adobe Photoshop® into 256 levels to provide lightness steps that are less than half the size of a ‘just notable difference’ (24). To determine the mean L^* value, we multiplied the Adobe Photoshop® histogram mean by 100 and then divided by 255. [L^* value mean = (Adobe Photoshop® histogram mean) (100)/255]. Developers of Adobe Photoshop® also limited the theoretically, unbounded range of a^* and b^* values to +127 to –127 to encode the color data without eliminating any but the most saturated self-luminous colors (24). Therefore a^* and b^* values are not scaled, but rather shifted so that an

a^* value of 128 is a CIELAB value of 0, and an a^* value of 138 is an CIELAB value of 10. We therefore transformed data for the lasso technique (the Adobe Photoshop® histogram mean a^* and b^* values) by using the following formulae: a^* value mean = Adobe Photoshop® histogram a^* mean–128; b^* value mean = Adobe Photoshop® histogram b^* mean–128.

Statistical analyses

Assessment of agreement of skin color metrics was conducted in two stages: 1) intra-rater and 2) inter-rater agreement. To evaluate agreement of repeated measurements of skin color metrics within individual raters (intra-rater agreement), a series of intra-class correlation coefficients (ICC) were computed using either 1) three observations from the lasso technique (skin region) or 2) nine observations from the color sampler technique (labia majora and vaginal wall regions). ICCs were computed using the *irr* package (31) of the R environment for statistical computing (32). For both the lasso and color sampler techniques, an ICC was computed for means recorded by each observer. Given the design of this study, ICCs for intra-rater agreement were specified as one-way random as observers were not randomly selected, but images were. Additionally, single values were specified as the unit of analysis (as opposed to averaged values) for these ICCs and agreement was computed in terms of consistency (rather than absolute agreement).

After computing intra-rater agreement, skin color values were averaged within each rater [within each anatomical region, for each color value (L^* , a^* , b^*), and for each technique (lasso, color sampler)]. Averaged values were then compared between raters (inter-rater agreement) using a series of one-way ICCs for consistency and average values.

Results

Results of ICCs for intra-rater agreement showed an extremely high level of internal consistency reliability within each rater for the lasso-technique and somewhat lower, yet acceptable, level of agreement for the color sampler technique (see Table 1). Both raters were equally consistent for skin areas, but Rater 1 (expert) was more consistent in ratings for the labia and vaginal wall than rater 2 (novice) (see Table 1).

Results of ICCs for inter-rater agreement were mixed across anatomical regions, with skin L^* , skin b^* , and labia L^* values reaching the highest level of agreement and skin a^* , labia b^* , and vaginal wall b^* possessing the lowest (see Table 2). Descriptive statistics for averaged skin color values from the expert rater are presented in Table 3 for reference and Fig. 2 portrays the scatterplots demonstrating inter-rater agreement graphically.

Discussion

When multiple raters perform DICA, accurate and precise data are needed to test hypotheses and maximize internal and external validity. We found that novice analysts can achieve high levels of agreement with expert analysts with training and the use of a detailed, standard protocol. The highest inter-rater reliability (ICCs = 0.93–0.98) occurred with L^* (light/dark) values of the skin and labia sites. As the L^* values are often used as the primary skin color variable (2, 4, 33), data describing the light/dark color measurement can be considered reproducible with both the lasso and color sampler technique. Color analysis of the vaginal wall, which is muscle tissue with many color variations, was less precise across both analysts [ICC = 0.82 for a^* (red/green) values and ICC = 0.74 for b^* (yellow/blue) values]. Lowest agreement (ICC = 0.64) occurred for the b^* values of the labia, an anatomical area with marked color variations as exemplified by a rather large standard deviation (SD = 13.67) of the b^* value measurements.

We sampled each labial and vaginal site nine times using an area of 25 pixels for each color sample (as compared to 120,000–150,000 pixels for the other skin sites: forearm, upper arm, thigh, sacrum, untanned sites). Overall, inter-rater reliability of data derived from the color sampler technique was less precise than those derived from the lasso technique. We propose several reasons for this finding. First, we noted more variation in color in the images of the labia and vagina as compared with the skin sites. Second, our sampling technique may need further refinement. The color sampler technique allowed us to collect four data points in a series before resetting the software and repeating the procedure. We gathered four data points, followed by four more data points, followed by a single final data point (4 + 4 + 1 samples). There were times when the representative color area was quite small; when selecting a series of four data points in a small area, we were forced to select a non-representative color sample as the fourth data point. We suspect that by collecting a series of three rather than four data points in a series, and repeating this procedure three times (3 + 3 + 3 samples), we would improve both intra- and inter-rater agreement by sampling a smaller and more uniform color area. Lower agreement in the color sampler technique (25 pixels) as compared with the lasso technique (120,000–150,000 pixels) suggests that, for small areas of skin color analysis, additional training may be required. Clear directions about the exact location for sampling (such as the use of the clock diagram to guide the selection of the image for analysis) are essential for reproducibility across analysts.

Additional training may be required for protocols identifying skin injury or pathology, which may result in significant alterations from normal in a^* and b^* values. Both intra-rater and inter-rater reliability were lower for a^* (red/green) and b^* (yellow/blue) values as compared with L^* values. For skin conditions such as mucosal tissue injury, erythema, and pressure ulcers, quantifying skin lightness/darkness only will not provide the detail needed to understand the extent or progression of the skin condition.

Limitations

There are several limitations to this study. Selection of the image folders for analysis was performed based on clarity and representativeness of skin color; no random assignment was used. While we attempted to include a range of skin colors in the images, there was no systematic process for selection. Image files were chosen by the expert analyst, and although the selection occurred prior to any image analysis, previous knowledge of the images may have biased the findings. Other errors that may have reduced the accuracy and precision of the findings included transcription and data entry errors, differences in the resolution and brightness of the laptop computer screens, and differences in ambient lighting during the analysis.

Conclusion

Digital image color analysis is a useful technique for determining skin color phenotype from digital images. Determination of skin color using multiple analysts reduces systemic bias that may occur from a single-color analyst. If multiple analysts are employed in a study protocol, however, they need training to ensure that intra-rater (repeatability) and inter-rater reliability (reproducibility) of the data occurs. We recommend using a standard protocol and an image testing procedure to determine intra- and inter-rater reliability prior to DICA required for a study protocol.

Acknowledgments

This work was supported by the following grants from the US National Institutes of Health (National Institute of Nursing Research and National Institute of Mental Health): Injury from Sexual Assault: Addressing Health

Disparity (2R01NR05352; Marilyn S. Sommers, PI), Injury in Latina Women after Sexual Assault: Moving Toward Health Care Equity (1R01NR011589; Marilyn S. Sommers, PI).

References

1. Parra EJ, Kittles RA, Shriver MD. Implications of correlations between skin color and genetic ancestry for biomedical research. *Nat Genet.* 2004; 36(Suppl 11):S54–S60. [PubMed: 15508005]
2. Visscher M, Sullivan D, Sullivan S, Barford B, Dock M, Sommers MS. Quantitation of epidermal and mucosal tissue injury using contrast agents and imaging techniques. *Skin Res Technol.* 2009; 15:180–186. [PubMed: 19622128]
3. Canning J, Barford B, Sullivan D, Wickett R, Visscher M. Use of digital photography and image analysis techniques to quantify erythema in health care workers. *Skin Res Technol.* 2009; 15:24–34. [PubMed: 19152575]
4. Sommers MS, Zink TM, Fargo JD, et al. Forensic sexual assault examination and genital injury: is skin color a source of health disparity? *Am J Emerg Med.* 2008; 26:857–866. [PubMed: 18926341]
5. Whitmer K, Barford B, Turner M, Sullivan D, Sommers M. Digital image analysis of facial erythema over time in persons with varied skin pigmentations. *Skin Res Technol.* 2011; 17:348–352. [PubMed: 21342293]
6. Kitzmiller WJ, Visscher M, Page DA, Wickett RR, Kitzmiller KW, Singer LJ. A controlled evaluation of dermabrasion versus CO2 laser resurfacing for the treatment of perioral wrinkles. *Plast Reconstr Surg.* 2000; 106:1366–1372. discussion 1373–1364. [PubMed: 11083571]
7. Clarys P, Alewaeters K, Lambracht R, Barel AO. Skin color measurements: comparison between three instruments: the Chromameter(R), the DermaSpectrometer(R) and the Mexameter(R). *Skin Res Technol.* 2000; 6:230–238. [PubMed: 11428962]
8. Takiwaki H. Measurement of skin color: practical application and theoretical considerations. *J Med Invest.* 1998; 44:121–126. [PubMed: 9597799]
9. Parra EJ. Human pigmentation variation: evolution, genetic basis, and implications for public health. *Am J Phys Anthropol.* 2007; 50:85–105. [PubMed: 18046745]
10. Imai FH, Tsumura T, Haneishi H, Miyake Y. Principal component analysis of skin color and its application to colorimetric color reproduction on CRT display and hardcopy. *J Imaging Sci Technol.* 1996; 40:422–430.
11. Parker KJ, Zhang M, Rubens DJ. An introduction to color in medical imaging. *J Imaging Sci Technol.* 2006; 50:12–16.
12. Ojima N, Akazaki S, Hori K, Tsumura N, Miyake Y. Application of image-based skin chromophore analysis to cosmetics. *J Imaging Sci Technol.* 2004; 48:222–226.
13. Preece, SJ.; Cotton, E.; Claridge, E. Imaging the pigments of human skin with a technique which is invariant to changes in surface geometry and intensity of illuminating light. In: Barber, D., editor. *Proceedings of Medical image understanding and analysis (MIUA).* England: Sheffield; 2003. p. 145-148.
14. Vezhnevets, V.; Sazonov, V.; Andreeva, A. A survey on pixel-based skin color detection techniques. Paper presented at: *Graphicon-2003*; Moscow, Russia. 2003.
15. Wan S, Parrish JA, Jaenicke KF. Quantitative evaluation of ultraviolet induced erythema. *Photochem Photobiol.* 1983; 37:643–648. [PubMed: 6611670]
16. Mattsson U, Cassuto J, Jontell M, Jonsson A, Sinclair R, Tarnow P. Digital image analysis of erythema development after experimental thermal injury to human skin: effect of postburn topical local anesthetics. *Anesth Analg.* 1999; 88:1131–1136. [PubMed: 10320183]
17. Aspres N, Egerton IB, Lim AC, Shumack SP. Imaging the skin. *Australas J Dermatol.* 2003; 44:19–27. [PubMed: 12581077]
18. Fairchild, MD. *Color appearance models.* Hoboken, NJ: Wiley; 2005.
19. Lee, H. *Introduction to color imaging science.* Cambridge, UK: Cambridge; 2005.
20. Reinhard, E.; Khan, E.; Akyuz, A.; Johnson, G. *Color imaging: fundamentals and applications.* Wellesley, MA: A.K. Peters, Ltd; 2008.
21. Berns, R. *Billmeyer and Saltzman's principles of color technology.* 3. New York: John Wiley & Sons; 2000.

22. Ha S, Lee M, Lee O, et al. A study of a method for distribution analysis of skin color. *Skin Res Technol.* 2009; 15:200–213. [PubMed: 19416467]
23. Alaluf S, Atkins D, Barrett K, Blount M, Carter N, Heath A. The impact of epidermal melanin on objective measurements of human skin colour. *Pigment Cell Res.* 2002; 15:119–126. [PubMed: 11936269]
24. Adobe Photoshop. TIFF technical notes. Mar 22. 2002 <http://partners.adobe.com/public/developer/en/tiff/TIFFphotoshop.pdf>
25. Chardon A, Cretois I, Hourseau C. Skin colour typology and suntanning pathways. *Int J Cosmet Sci.* 1991; 13:191–208. [PubMed: 19291061]
26. Sommers MS, Fargo JD, Baker RB, Fisher B, Buschur C, Zink TM. Health disparities in the forensic sexual assault examination related to skin color. *J Forensic Nurs.* 2009; 5:191–200. [PubMed: 19947958]
27. PictoColor inCamera Software. 2013. <http://www.pictocolor.com/incamera.htm>
28. Munsell ColorChecker Mini Card. 2008; 2013 <http://www.imagescience.com.au/products/X%252dRite-Munsell-ColorChecker-Mini.html>.
29. Rosvall MD, Fields HW, Ziuchkovski J, Rosenstiel SF, Johnston WM. Attractiveness, acceptability, and value of orthodontic appliances. *Am J Orthod Dentofac Orthop.* 2009; 135:276–277.
30. Stevens M, Stoddard M, Higham J. Studying primate color: towards visual system-dependent methods. *Int J Primatol.* 2009; 30:893–917.
31. Gamer, M.; Lemon, J.; Singh, I. irr: various coefficients of interrater reliability and agreement. R package. Version 0.84. 2012. <http://CRAN.R-project.org/package=irr>
32. Team RDC. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012.
33. Baker RB, Sommers MS. Relationship of genital injuries and age in adolescent and young adult rape survivors. *J Obstet Gynecol Neonatal Nurs.* 2008; 37:282–289.

Fig. 1. Screenshot of Munsell® Color Checker mini card 28 using Pictocolor® in Camera software²⁷ for digital image color correction.

Fig. 2.

Scatterplots of L^* , a^* , and b^* color values for skin, labia, and vaginal wall anatomical regions from both raters.

TABLE 1

Results (mean) of intra-rater agreement of skin color metrics

	Rater 1 (expert)	Rater 2 (novice)
Skin		
L^*M	1.00	1.00
a^*M	1.00	1.00
b^*M	1.00	1.00
Labia		
L^*M	0.99	0.88
a^*M	0.96	0.89
b^*M	0.98	0.81
Vaginal wall		
L^*M	0.98	0.98
a^*M	0.94	0.86
b^*M	0.91	0.88

Values are one-way intra-class correlation coefficients M = Mean.

TABLE 2

Results of inter-rater agreement of skin color metrics

	Skin	Labia	Vaginal wall
L^*	0.98	0.93	0.83
a^*	0.76	0.83	0.82
b^*	0.92	0.64	0.74

Values are one-way intra-class correlation coefficients.

TABLE 3

Descriptive statistics for skin color metrics for expert rater

Anatomical region	M	SD	Min	Max
Skin (lasso)				
L^*	45.07	16.55	19.33	84.14
a^*	12.86	5.45	2.66	30.42
b^*	22.12	6.76	8.24	37.44
Labia (color sampler)				
L^*	49.67	10.35	35.44	67.22
a^*	27.74	9.45	15.44	45.33
b^*	17.52	13.67	5.44	53.56
Vaginal wall (color sampler)				
L^*	56.39	9.84	38.11	73.44
a^*	24.05	9.27	6.11	41.22
b^*	9.58	5.48	2.00	19.22

M = Mean; SD = Standard Deviation; Min = Minimum Value; Max = Maximum Value.