ORIGINAL ARTICLE

https://doi.org/10.1590/1806-9282.20210982

Intra- and inter-examiner reliability of digital images of skin donor areas in burns

Flávia Fernanda de Oliveira Assunção[1], Rosana Caetano Gomes[1], Nathália Cristina de Souza Borges[1], Luiz Otávio Murta Junior[2], Jayme Adriano Farina Júnior[3], Elaine Caldeira de Oliveira Guirro[1,4], Rinaldo Roberto de Jesus Guirro[1,4*]

SUMMARY

OBJECTIVE: The main objective of this study was to evaluate intra- and inter-rater reliability in the analysis of digital images of donor areas for skin in burn patients using the CaPAS plugin in the ImageJ®.

METHODS: Donor sites were reviewed by two independent reviewers in duplicate. The capture of images was standardized on the same device and distance (with a millimeter ruler), without a flash. The evaluators were trained to capture the images and use the plugin.

RESULTS: We selected 70 images from donor areas, from men and women between 18 and 60 years old. In the analysis of intra-examiner reliability, eight of the nine variables exhibited excellent reliability (0.985–0.998) and one (entropy) exhibited good reliability (0.525). The same was true for the inter-examiner analysis: excellent reliability for eight variables (0.824–0.993) and good reliability for entropy (0.501).

CONCLUSIONS: The CaPAS plugin has proven to be a reliable tool for use in research in skin donor areas in burns, as demonstrated by its excellent intra- and inter-examiner reliability values. This is a pioneering study in the quantitative assessment of skin donor areas in burn patients using the CaPAS plugin.

KEYWORDS: Reproducibility of results. Data accuracy. Burns.

INTRODUCTION

Digital photographs have been used for skin analysis in research with an interest in more reliable investigations of tissue characteristics[1]. Creating new tools to quantify and validate these images is a challenge[2], since a subjective analysis only provides qualitative data that are difficult to reproduce[3]. The clinical evaluation of the burned patient (both donor and recipient areas) is routinely performed qualitatively through direct inspection[4].

The easy-to-use ImageJ® public domain software developed in Java programming at the U.S. National Institutes of Health is a well-known health tool for image processing that has also been used for the assessment of skin lesions and can, therefore, be very useful in the evaluation of burn patients[5].

The CaPAS plugin was designed for carotid plaque analysis and was developed for the quantitative evaluation of videodensiometric images of atherosclerosis through computational measurements. This tool is sensitive to grayscale changes in cholesterol, collagen, and calcium, enabling the differentiation of groups[6], patients with and without symptoms have a different plaque composition[7].

The use of the CaPAS plugin in burn patients was reported in a study that analyzed images of skin donor areas (e.g., thigh and scalp) in two groups (i.e., sham and intervention) to investigate re-epithelialization[8]. The plugin made it possible to quantify the images and differentiate the texture of the re-epithelialized tissue. However, no plugin properties or variables were evaluated.

The aim of this study was to evaluate intra- and inter-rater reliability in the analysis of images of skin donor areas in burn patients using the CaPAS plugin implemented in ImageJ®.

METHODS

A total of 70 images (32 scalp and 38 thigh) of 70 patients were taken from skin donor areas of burn patients admitted to the emergency unit of the hospital of the Ribeirão Preto School of Medicine. One was randomly selected using a simple lottery

[1]Universidade de São Paulo, Ribeirão Preto Medical School, Postgraduate Program in Functional Rehabilitation and Performance – Ribeirão Preto (SP), Brazil.
[2]Universidade de São Paulo, Ribeirão Preto School of Philosophy, Sciences and Letters, Department of Physics and Mathematics – Ribeirão Preto (SP), Brazil.
[3]Universidade de São Paulo, Ribeirão Preto Medical School, Department of Surgery – Ribeirão Preto (SP), Brazil.
[4]Universidade de São Paulo, Ribeirão Preto Medical School, Department of Health Sciences – Ribeirão Preto (SP), Brazil.
*Corresponding author: rguirro@fmrp.usp.br
Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on October 13, 2021. Accepted on January 06, 2022.

system for analysis by two evaluators, according to previous health reliability studies[9].

Images were taken using a 12-megapixel digital camera (DSC-W510, Sony, Manaus, Brazil) with 72 dpi. The images were taken without flash with the camera perpendicular to the area to be photographed at a standardized distance of 30 cm apart (determined using a millimeter ruler) on the same plane as the sample. This was performed by two independent evaluators who had undergone training for the use of the software. All images were analyzed twice, with a 1-week interval between evaluations. The evaluation and digitization of the analyses followed the same order (1–70) for both evaluators. The image evaluation procedure is described in Figure 1.

Outline the area of interest with the computer mouse. Select the ANALYSIS icon. Then, the graph with the nine measurements will be generated: Count: number of pixels within the area; Average: average pixel value calculated on the pixels in the selected region; Standard Deviation: of the gray values of the pixel in the selected region; Skewness: asymmetry (when positive, asymmetry is toward the dark side; when negative, asymmetry is toward the light side); Kurtosis: graph symmetry, Energy: angular value, when the elements of the co-occurrence matrix are very unequal; Inertia: intensity contrast between a pixel and its neighbor; Entropy: degree of uncertainty in the uniformity of the selected region (its value increases when homogeneity reduces); and Homogeneity: dissimilarity and contrast.

These measurements are represented by texture parameters involving mean and standard deviation, and grayscale parameters (i.e., entropy, energy, and homogeneity). The mean gray level (MGL) represents the average grayscale value of the pixels in the region of interest on a scale of 256 shades of gray, with zero being the darkest possible tone (hypoechoic) and 255 the lightest possible tone (hyperechoic)[6].

Intra- and inter-rater reliability were determined using the intraclass correlation coefficient (ICC2,1), with the calculation of the 95% confidence interval, standard error of measurement (SEM), and minimal detectable change (MDC). ICC values were interpreted based on Weir[10]: 1.00–0.81 indicates excellent reliability; 0.80–0.61 indicates very good reliability; 0.60–0.41 indicates good reliability; 0.40–0.21 indicates reasonable reliability; and 0.20–0.00 indicates poor reliability. The statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS), version 17.0 (Chicago, IL, USA).

RESULTS

A total of 70 digital photographic images of 70 patients of both sexes between the ages of 18 and 60 years were evaluated. The mean values are described in Table 1.

Table 2 shows intra- and inter-rater ICC values. Intra-rater ICC values ranged from 0.985 to 0.998 (excellent reliability) for eight of the nine variables. The intra-rater ICC for "entropy" was 0.525 and inter-rater ICC for "entropy" was 0.501, thus good reliability was found for entropy, while excellent reliability was found for other variables (ICC=0.824–0.993).

DISCUSSION

The evaluation of burns is commonly performed by local visual inspection, being a quick and easy method, but subjective, not being the most reliable or accurate[11,12], as it strongly depends on the experience of the evaluator.

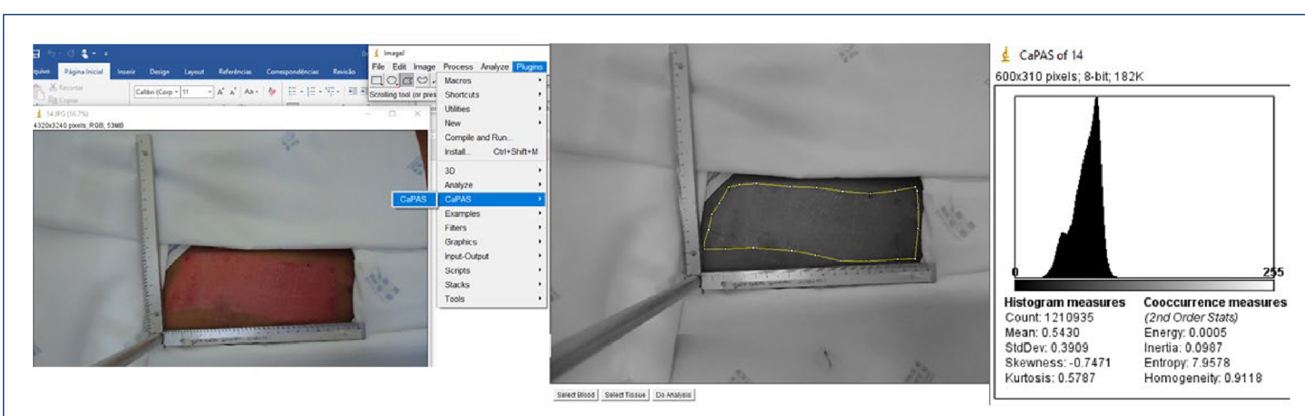

Figure 1. Importing the image in the File icon. After inserting the image, click twice on the CaPAS icon. At this moment, the image will appear in grayscale.

Table 1. Description of the mean of the variables by both evaluators.

Variables	Evaluator 1		Evaluator 2	
	Test	Retest	Test	Retest
Count	490,280 (361,584)	480,188 (369,297)	447,302 (342,365)	453,803 (349,364)
Mean	0.73 (0.23)	0.73 (0.22)	0.72 (0.23)	0.72 (0.23)
Standard deviation	0.53 (0.16)	0.52 (0.16)	0.52 (0.17)	0.52 (0.16)
Skewness	-0.76 (0.03)	-0.76 (0.02)	-0.75 (0.05)	-0.76 (0.03)
Kurtosis	0.60 (0.05)	0.60 (0.05)	0.62 (0.16)	0.60 (0.05)
Energy	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Inertia	0.04 (0.03)	0.04 (0.03)	0.05 (0.07)	0.05 (0.04)
Entropy	8.99 (0.60)	9.00 (0.61)	8.97 (0.64)	8.86 (1.18)
Homogeneity	0.95 (0.03)	0.95 (0.03)	0.95 (0.03)	0.95 (0.03)

Data expressed as mean (standard deviation).

Table 2. Summary of intra- and inter-examiner reliability obtained.

Variables	ICC	95%CI	SEM (%)	MDC (pixels)
Intra-rater				
Count	0.997	0.995–0.998	4.13	55,481.56
Mean	0.998	0.996–0.998	1.40	0.03
Standard deviation	0.997	0.996–0.998	1.66	0.02
Skewness	0.991	0.986–0.995	0.38	0.01
Kurtosis	0.992	0.987–0.995	0.86	0.01
Energy	0.989	0.982–0.993	8.93	0.00
Inertia	0.980	0.967–0.987	11.28	0.02
Entropy	0.525	0.236–0.705	4.67	1.17
Homogeneity	0.985	0.976–0.991	0.41	0.01
Inter-rater				
Count	0.983	0.972–0.989	10.03	129,864.35
Mean	0.993	0.989–0.996	2.68	0.05
Standard deviation	0.993	0.988–0.995	2.59	0.04
Skewness	0.985	0.975–0.990	0.48	0.01
Kurtosis	0.987	0.980–0.992	1.09	0.02
Energy	0.824	0.717–0.891	35.53	0.00
Inertia	0.895	0.831–0.935	26.35	0.04
Entropy	0.501	0.197–0.690	7.11	1.76
Homogeneity	0.903	0.843–0.939	1.06	0.03

ICC: intraclass correlation coefficient; CI: confidence interval; SEM: standard error of measurement (% score); MDC: minimal detectable change (% score).

The use of digital photographs is a viable alternative to overcome this limitation. It has been used in research on skin erythema[13], skin cancer[14], and vocal folds[15], seeking to quantify differences between treatments and/or textures, and is also used in clinical practice.

Skin assessment can be performed using instruments such as the Vancouver Scar Scale and the Patient and Observer Scar Assessment Scale[16], which require training and specific knowledge of their psychometric properties. Another possibility is the laser Doppler, which is expensive and not always available in burn centers[17].

Hop et al.[18] highlighted the need for more accurate assessments of burns, as the cost of hospitalizations and surgeries is very high and optimal treatment requires an accurate diagnosis. As demonstrated in this study, the use of quantitative assessment based on digital photographs is highly reproducible and meets the requirements for assessing skin donor areas for grafts in burn patients.

As this is a reliability study, whose objective was to determine the reproducibility of using the CaPAS plugin as a tool to assess the skin surface, the images used were standardized in a skin donor area. The removal of the skin was always performed with the same thickness, dermatome, and defined edges, allowing to differentiate the intact skin. Excellent intra- and inter-examiner reliability was found for eight of the nine variables of the CaPAS plugin and very good reliability for the variable "entropy." The lower reliability of this variable may be due to its greater sensitivity when performed by different evaluators, as it depends on the demarcation of the image and the grayscale pixel values. With lighter or darker pixels, an important difference in variables can occur even with a difference in demarcation of millimeters between evaluators.

In a study that investigated the re-epithelialization of the donor area in burns, the inertia and homogeneity variables were able to differentiate different groups[8]. In this research, it is noteworthy, therefore, that both variables showed excellent intra- and inter-observer correlation for the same population, making their use reliable and reproducible as a quantitative way of evaluating the donor area.

Reliability studies consider ICC values >0.40 acceptable[19]. In this research, the intra- and inter-examiner ICC values found are all acceptable for reliability. Originally colored images are transformed into black and white during the plugin execution, which may explain the small divergence between evaluators regarding the demarcation of the selected area's borders. The good intra-examiner correlation for this item validates those of the plugin, as the intensity of a pixel is more important than its homogeneity.

The statistical value of energy increases when the co-occurring matrix elements are very unequal, which means large differences in texture patterns within the skin removal area. Homogeneity measurements reflect the level of roughness of the donor area. Thus, when homogeneity is high, the homogeneous distribution of texture patterns is found within the selected skin donor area, which means irregular but homogeneous pixel tones.

The high "count" values, which correspond to the pixel area, obtained in the intra- and inter-rater analyses result from the calculations performed, in which the area is multiplied by millimeters squared (mm^2) of pixels, which requires the use of a millimeter ruler on the same plane as the area of interest when the digital image is captured.

In reliability studies, it is important to establish the amount of error inherent to a given measurement method[20]. The SEM indicates errors when measuring a particular variable with an assessment tool, i.e., it is an indication of the accuracy of a score. Regarding the MDC, Weir[10] emphasized that it is important to be aware of the minimum difference in scores of an assessment instrument between reevaluations.

The present study offers a promising analysis of the quantitative evaluation of skin donor areas in burn patients, capable of detecting small variations in the skin surface. This method is reliable and allows for more rigorous evaluation of these patients. In fact, most studies involving the evaluation of burned skin or donor area involve qualitative analysis, whose results may be imprecise, as they depend on the evaluator's perception, experience, and criteria. The analysis of digital photographs using the CaPAS plugin also allows the monitoring of the healing process and the effect of different treatments in both research and clinical practice.

The reliability of this tool allows its use in this population, as well as in other populations in which skin repair is investigated. ImageJ® software is easy-to-use and public domain, so it is free and only a digital photograph is required to run the plugin, in addition to providing fast results, making it an extremely valuable assessment tool for burn patients.

CONCLUSION

The CaPAS plugin implemented in ImageJ® software proved to be a reliable tool for use in research and clinical practice involving the analysis of skin donor areas in burn victims, as demonstrated by the excellent intra- and inter-rater reliability values.

AUTHORS' CONTRIBUTIONS

FFOA: Conceptualization, Investigation, Formal Analysis, Supervision, Writing – original draft, Validation, Visualization, Project administration, Writing – review & editing. **RCG:** Investigation, Formal Analysis, Writing – original draft, Validation, Visualization, Project administration, Writing – review & editing. **NCSB:** Data curation. **LOMJ:** Methodology, Resources, Software, Validation, Visualization, Project administration, Writing – review & editing. **JAFJ:** Methodology, Validation, Visualization, Project administration, Writing – review & editing. **ECOG:** Supervision, Validation, Visualization, Project administration, Writing – review & editing. **RRJG:** Conceptualization, Formal Analysis, Funding acquisition, Methodology, Supervision, Validation, Visualization, Project administration, Writing – review & editing.

REFERENCES

1. Choi JY, Kim EJ, Jang SI, Kim AR, Lee TJ, Lee HK. A new technique for evaluating heel xerosis grade and the effects of moisturizer on heel skin dryness. Skin Res Technol. 2018;24(4):557-61. https://doi.org/10.1111/srt.12466
2. Li D, Mathews C, Zhang F. The characteristics of pressure injury photographs from the electronic health record in clinical settings. J Clin Nurs. 2018;27(3-4):819-28. https://doi.org/10.1111/jocn.14124
3. Sabetai MM, Tegos TJ, Nicolaides AN, Dhanjil S, Pare GJ, Stevens JM. Reproducibility of computer-quantified carotid plaque echogenicity: can we overcome the subjectivity? Stroke. 2000;31(9):2189-96. https://doi.org/10.1161/01.str.31.9.2189
4. Nascimento C, Hansen LD, Sandoval ML, Santos VN, Vieira ALN, Ramos FS. Treatment of sequelae of burn: a case study. Rev Bras Queimaduras. 2014;13(4):267-70.
5. Rodrigues DF, Mendes FF, Menezes LB, Carvalho WL, Sá S, Silva JA, et al. Tratamento de feridas excisionais de coelhos com extrato de barbatimão associado a células mononucleares autólogas da medula óssea. Arq. Bras. Med. Vet. Zootec. 2017;69(5):1243-50. https://doi.org/10.1590/1678-4162-9301
6. Baroncini LA, Pazin Filho A, Murta Junior LO, Martins AR, Ramos SG, Cherri J, et al. Ultrasonic tissue characterization of vulnerable carotid plaque: correlation between video densitometric method and histological examination. Cardiovasc Ultrasound. 2006;4:32. https://doi.org/10.1186/1476-7120-4-32
7. Nandalur KR, Baskurt E, Hagspiel KD, Phillips CD, Kramer CM. Calcified carotid atherosclerotic plaque is associated less with ischemic symptoms than is noncalcified plaque on MDCT. AJR Am J Roentgenol. 2005;184(1):295-8. https://doi.org/10.2214/ajr.184.1.01840295
8. Gomes RC, Guirro ECO, Gonçalves AC, Farina Junior JA, Murta Junior LO, Guirro RRJ. High-voltage electric stimulation of the donor site of skin grafts accelerates the healing process. A randomized blinded clinical trial. Burns. 2018;44(3):636-45. https://doi.org/10.1016/j.burns.2017.10.006
9. Bassi D, Santos-de-Araújo AD, Camargo PF, Dibai-Filho AV, Fonseca MA, Mendes RG, et al. Inter and intra-rater reliability of short-term measurement of heart rate variability on rest in diabetic type 2 patients. J Med Syst. 2018;42(12):236. https://doi.org/10.1007/s10916-018-1101-8
10. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231-40. https://doi.org/10.1519/15184.1
11. Pape SA, Skouras CA, Byrne PO. An audit of the use of laser Doppler imaging (LDI) in the assessment of burns of intermediate depth. Burns. 2001;27:233-9. https://doi.org/10.1016/s0305-4179(00)00118-2
12. Droog EJ, Steenbergen W, Sjöberg F. Measurement of depth of burns by laser Doppler perfusion imaging. Burns. 2001;27(6):561-8. https://doi.org/10.1016/s0305-4179(01)00021-3
13. Cho M, Lee DH, Doh EJ, Kim Y, Chung JH, Kim HC, et al. Development and clinical validation of a novel photography-based skin erythema evaluation system: a comparison with the calculated consensus of dermatologists. Int J Cosmet Sci. 2017;39(4):426-34. https://doi.org/10.1111/ics.12393
14. Rosenberg A, Meyerle JH. Total-body photography in skin cancer screening: the clinical utility of standardized imaging. Cutis. 2017;99(5):312-6. PMID: 28632800
15. Song CI, Ryu CH, Choi SH, Roh JL, Nam SY, Kim SY. Quantitative evaluation of vocal-fold mucosal irregularities using GLCM-based texture analysis. Laryngoscope. 2013;123(11):E45-50. https://doi.org/10.1002/lary.24151
16. Schiefer JL, Rath R, Ahrens E, Grigutsch D, Gräff I, Stromps JP, et al. Evaluation of scar quality after treatment of superficial burns of the hands and face with Dressilk or Biobrane — An intra-individual comparison. Burns. 2018;44(2):305-17. https://doi.org/10.1016/j.burns.2017.07.024
17. Alsharnoubi J, Mohamed O. Photobiomodulation effect on children's scars. Lasers Med Sci. 2018;33(3):497-501. https://doi.org/10.1007/s10103-017-2387-3.
18. Hop MJ, Stekelenburg CM, Hiddingh J, Kuipers HC, Middelkoop E, Nieuwenhuis MK, et al. Cost-effectiveness of laser doppler imaging in burn care in the Netherlands: a randomized controlled trial. Plast Reconstr Surg. 2016;137(1):166e-176e. https://doi.org/10.1097/PRS.0000000000001900.
19. Guirro ECO, Leite GPMF, Dibai-Filho AV, Borges NCS, Guirro RRJ. Intra and inter-rater reliability of peripheral arterial blood flow velocity by means of doppler ultrasound. J Manipulative Physiol Ther. 2017;40(4):236-40. https://doi.org/10.1016/j.jmpt.2017.02.007
20. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96-106. https://doi.org/10.1016/j.jclinepi.2010.03.002