Original Article

DOI: 10.5582/ddt.2025.01104

Three-dimensional imaging provides reliable size measurement of skin lesions associated with vascular anomalies: A comparison with two-dimensional photography

Kana Sakai[1], Kayo Kunimoto[1,*], Yuna Noda[1], Yutaka Inaba[1], Yuki Yamamoto[1], Shiho Yasue[2], Akifumi Nozawa[2], Daichi Hayashi[2], Saori Endo[2], Michio Ozeki[2], Masatoshi Jinnin[1]

[1] Department of Dermatology, Wakayama Medical University, Wakayama, Japan;
[2] Department of Pediatrics, Graduate School of Medicine, Gifu University, Gifu, Japan.

SUMMARY: Three-dimensional (3D) imaging techniques enable accurate quantitative size measurement, and have been used to evaluate treatment effects on skin lesions such as ulcers, burns, and skin laxity. This study aimed to establish and validate a 3D imaging–based method to evaluate the surface area, area, and volume of cutaneous lesions associated with vascular anomalies (VA). We compared measurements obtained from two-dimensional (2D) photographs traced by three dermatologists with those obtained from 3D images traced by three company operators, and assessed inter- and intra-rater reliability. The procedure in the present study involves tracing lesion contours using photographs of VA captured by a 3D camera, followed by 3D processing of the images to measure lesion area, surface area, and volume. All patients provided written informed consent, and the study protocol was approved by the institutional review boards. Both 2D and 3D methods demonstrated high inter- and intra-rater reliability; however, better reliability was observed in the measurements obtained by company operators using 3D imaging. The findings indicate that 3D surface imaging provides more consistent and objective evaluation of lesion size than 2D photography and support the potential application of this method in clinical practice and clinical trials for VA. Accurate quantitative measurement of lesion size as an endpoint may facilitate the development of new treatment options for patients with VA.

Keywords: vascular anomalies, three-dimensional, surface area, area, volume

1. Introduction

Vascular anomaly (VA) is a general term for diseases caused by abnormal formation of blood vessels and/or lymphatic vessels. According to the International Society for the Study of Vascular Anomalies (ISSVA) classification, diseases of VA are divided into two major groups: vascular tumors and vascular malformations (*1*). Vascular tumors are further classified into three categories: benign, locally invasive/borderline, and malignant, and each category includes multiple diseases (*e.g.* infantile hemangioma and tufted angioma [TA] of benign type) (*2*). On the other hand, vascular malformations are classified into "simple" "combined" "of major named vessels" and "associated with other anomalies". Simple vascular malformations are subclassified into venous malformation (VM), lymphatic malformation (LM), capillary malformation, and arteriovenous malformation.

VA occurs in tissues throughout the body, but are particularly common on the body surface and subcutaneous tissue, which manifest a variety of clinical presentations ranging from superficial macules to deep masses. Because these lesions are congenital or occur in childhood and afflict patients throughout their lives, there is a need to develop new treatments. Although there is no consensus measure for VA, a reduction of lesion volume ≥ 20% is commonly used as a definition of therapeutic response (*3-5*). Therefore, capturing the change in size (area or volume) is considered to be one of the most important indicators for evaluating drug efficacy in clinical trials for skin lesions associated with VA. However, in the case of skin lesions, different from visceral tumors, it is technically difficult to evaluate lesion volume accurately by CT, MRI or ultrasound, because these simple measurements include many normal areas due to the complex shape with irregularities.

Measurement techniques for surface area as well as area and volume using 3D images have been shown to be highly accurate and precise, and have been used

for the evaluation of treatment effects on skin lesions such as ulcers, burns and, laxity (*6-8*). Therefore, in this study, we aimed to establish a method to evaluate the size (surface area, area, and volume) of skin lesions of VA traced by multiple raters on 2D photographs or 3D images, *via* assessing the inter-rater and intra-rater reliability.

2. Materials and Methods

2.1. Ethical consideration

The present study was approved by the institutional review board of each participating institute (No. 4209). All patients voluntarily participated in this study and provided written informed consent. Legal representatives (*e.g.*, parent or guardian) provided consent on behalf of children (under 18 years of age at the time of obtaining consent).

2.2. Inclusion criteria

Patients who fulfill all of the following criteria were eligible for participation in the study: 1) Patients with skin lesions of VA in Wakayama Medical University Hospital and Gifu University Hospital; 2) Patients with at least one measurable target lesion of 2-420 cm^2 in area that can be assessed photographically using QuantifiCare LifeViz® Mini 3D camera (*6-8*) (QuantifiCare Inc., Cumming, GA); 3) Patients who agree to have their skin lesions photographed.

2.3. Exclusion criteria

Patients were excluded from participation in the study if they have any of the following: 1) Patients with significant bleeding in the target lesion or surrounding area; 2) Patients who were judged by the principal investigator or others to be unsuitable for enrollment in the study.

2.4. Study procedure

1) Two-dimensional (2D) photographs of the target lesions were obtained using QuantifiCare LifeViz® Mini 3D camera. Photographs for analysis should be as follows (Figure 1): i. The entire outline of the target lesion can be seen in a single photograph; ii. The target lesion is located on a flat area of the body; iii. The target lesion is photographed from the front; iv. The target lesion is not covered by hair; v. The photograph is not blurred, unclear, *etc.*; vi. A calibration sticker was applied near the target lesion.

2) The 2D photographs (Figure 2) were sent to Nobelpharma Co., Ltd. (Shinkawa, Tokyo, Japan) with clinical information (*e.g.* diagnosis, age, gender, and lesion site). In addition, a copy of each photograph with the target lesion marked was also provided (Figure 2b). Nobelpharma forwarded these image files to QuantifiCare Inc.

3a) Quantificare Inc. uploaded the 2D photographs (stereo images) on the online server (Dermapix Web Application version 1.8.10) for tracing the lesion

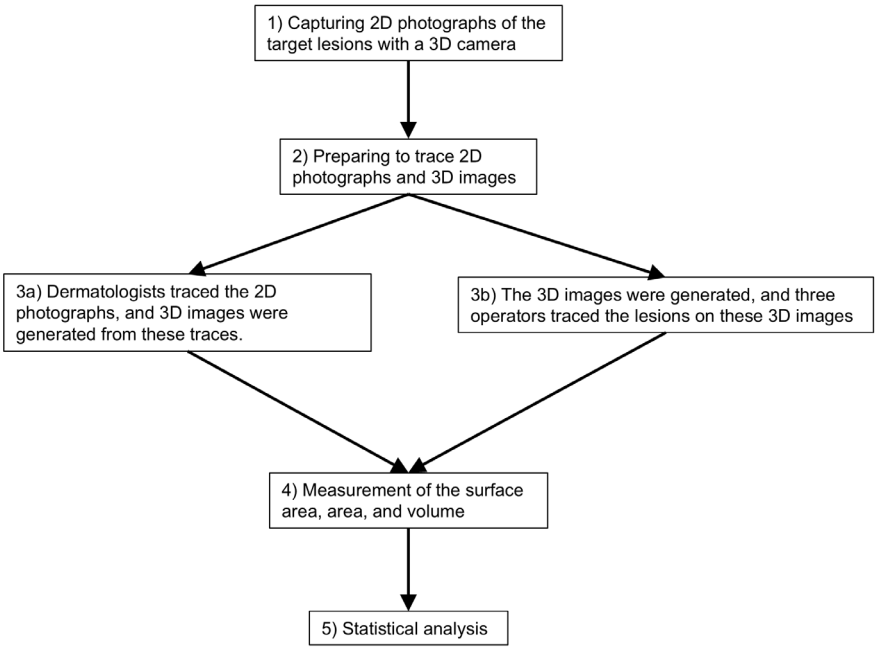

Figure 1. Scheme of study procedure. 1). Capture of 2D photographs of the target lesions using a 3D camera. 2). Preparation for tracing the 2D photographs and 3D images. 3a). Dermatologists traced the 2D photographs, and 3D images were generated from these traces. 3b). The 3D images were generated, and three operators traced the lesions on these 3D images. 4). Measurement of the surface area, area, and volume from the processed images. 5). Statistical analysis of the obtained measurements.

contours. Three dermatologists independent from the study each traced the lesion contours on 2D photographs of the online server (Figure 2c). To evaluate reproducibility, the trace was performed twice with a two-day interval. Quantificare Inc. created 3D images from the traced 2D photographs using QuantifiCare Suite (including DermaPix Software, Stitching, 3D reconstruction and 3D Analysis/3D Track, Version:3.16.4) (Figure 2d).

3b) Quantificare Inc. created 3D images from the untraced 2D photographs using QuantifiCare Suite (Figure 2e) and rendered the 3D image color-coded by depth. Three operators of Quantificare Inc., who were not involved in the study, also traced the lesion contours twice, by referring to color-coded 3D images (Figure 2f). This tracing was also projected onto the 2D photograph (Figure 2g).

4) Quantificare Inc. measured the surface area and volume on the traced 3D images, and measured the area from the traced 2D photographs. Area of 2D photographs was calculated using the scale provided by the calibration sticker. The 3D image was calibrated by combining the images of the calibration stickers to confirm the geometry of the 3D object. A fine grid was set up on the 3D constructed image, and the surface area of the lesion that completely covers the grid frame was calculated. Volume was also calculated based on the 3D images.

5) Nobelpharma Co., Ltd. performed statistical analysis as described below.

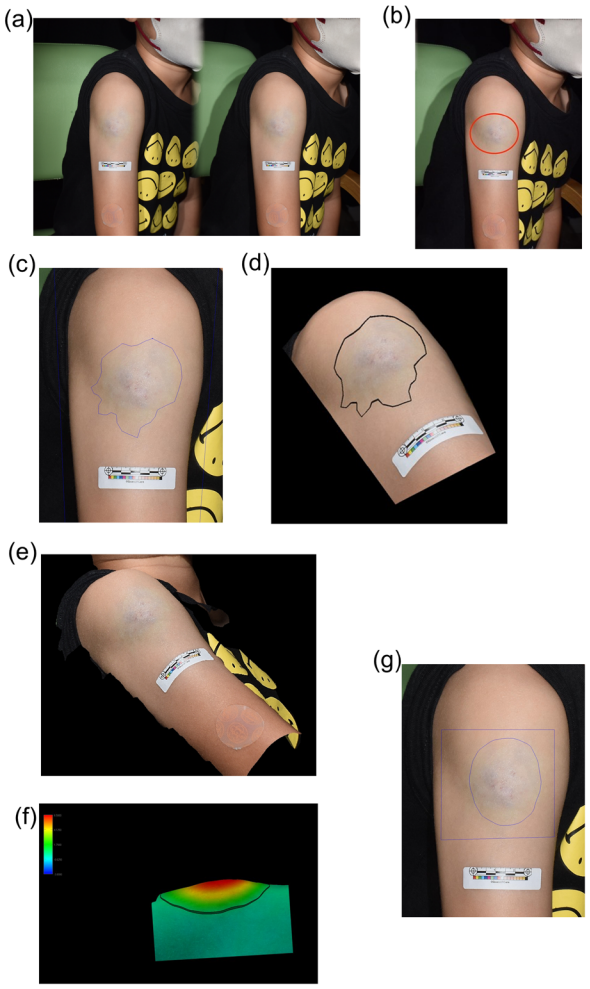

Figure 2. 3D processing of the images of patients with vascular anomalies. (a) An example of a 2D photograph (stereo image) of a patient with venous malformation (VM) of the right upper arm (case A1) obtained by the 3D camera. A calibration sticker was applied near the lesion. **(b)** An example of a 2D photograph in which the target lesion from Figure 2a was marked with a red circle by an investigator. **(c)** An example of the 2D photograph from Figure 2a traced by dermatologist rater 1. **(d)** An example of a 3D image reflecting the trace performed by dermatologist rater 1 as shown in Figure 2c. **(e)** An example of a 3D image processed using the 2D photograph from Figure 2a. **(f)** An example of 3D images color-coded by depth and traced by QuantifiCare operator 1. **(g)** An example of a 2D photograph reflecting the trace performed by QuantifiCare operator 4 as shown in Figure 2f.

2.5. Statistical analysis

1) The study analyses were performed as follows: i) Summary statistics (mean and geometric mean, standard deviation [SD] and geometric SD, and coefficient of variation [CV]%) for each parameter (surface area, area, and volume); ii) Intra-rater reliability (intra-class correlation coefficient [ICC]: intra-rater agreement); iii) Inter-rater reliability (ICC: inter-rater agreement); iv) Inter-rater reliability considering repeated assessments (ICC: inter-rater consistency).

Surface area was analysed as the primary endpoint using the following method. Area and volume were analysed in the same manner as exploratory endpoints.

Let Xijk denoted the k-th measurement of the surface area, traced by the j-th rater on the i-th lesion, where i = 1, 2, 3, ...n, j = 1, 2, 3, k = 1, 2. The repeated-measures model for reliability study with equal numbers in the subclass (*i.e.*, no missing data) was defined as

$$Xijk = \mu + \pi i + \gamma j + (\pi\gamma)ij + \varepsilon ijk$$

where μ was the overall population mean of measurements and πi and γj, are the subject and rater effects, respectively. The terms $(\pi\gamma)ij$ and εijk represented the inter-rater and intra-rater random errors. The components πi, and εijk were assumed to vary normally with means of zero and variances of σL^2 and σe^2, respectively; they were independent of each other and of all other components in the model. If the rater effects were random, then the components γj, and $(\pi\gamma)ij$ were assumed to vary normally with means of zero and variances of σR^2, and σLR^2, respectively; they were also independent of each other and of all other components.

2) Partial intra-rater reliability For each rater, intraclass correlation coefficients (ICC) (1, 1) was estimated based on a one-way ANOVA model. ($Xik = \mu + \pi i + \varepsilon ik$) as ICC (1, 1) = $\sigma L^2 / (\sigma L^2 + \sigma e^2)$, and the standard error (SE) was calculated as sqrt (σe^2). The 95%CI of ICC (1, 1) was also computed according to Shrout's method (*9*).

3) Partial inter-rater reliability For each measurement (*i.e.*, 1st or 2nd), ICC (2, 1) was estimated based on a two-way ANOVA model ($Xij = \mu + \pi i + \gamma j + \varepsilon ij$) as $ICC(2,1) = \sigma L^2/(\sigma L^2 + \sigma R^2 + \sigma e^2)$, and the SE was calculated as $SE = sqrt(\sigma R^2 + \sigma e^2)$.

The 95%CI of ICC (2, 1) was also computed according to Shrout's method (*9*).

4) Overall Inter-rater and intra-rater reliability with repeated measurements Based on the repeated two-way ANOVA model as described above, the inter-rater reliability coefficient (ρinter) and intra-rater reliability coefficient (ρintra) were estimated as $\rho inter = \sigma L^2 / (\sigma L^2 + \sigma R^2 + \sigma LR^2 + \sigma e^2)$, and $\rho intra = (\sigma L^2 + \sigma R^2 + \sigma LR^2) / (\sigma L^2 + \sigma R^2 + \sigma LR^2 + \sigma e^2)$ respectively. The standard error (SE) was calculated as $SEinter = sqrt(\sigma R^2 + \sigma LR^2 + \sigma e^2)$, and $SEintra = sqrt(\sigma e^2)$. The 95%CI of ρinter and ρintra was computed according to Shrout's method (*9*).

5) Interpretation of ICC The interpretation of ICC according to Fleiss was typically categorized as follows: < 0 = Less than chance agreement, 0.01-0.20 = Slight agreement; 0.21-0.40 = Fair agreement, 0.41-0.60 = Moderate agreement, 0.61-0.80 = Substantial agreement, and 0.81-0.99 = Almost perfect agreement (*10*).

All analyses were performed in SAS version 9.4.

3. Results

3.1. Participant demographics

In the present study, 16 lesions in 9 patients with VM (n = 5), LM (n = 3), or TA (n = 1) were included: The number of subjects, lesions measured, diagnosis, gender, and lesion site are summarized in Supplementary Table S1 (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=287*). In 2 patients, photographs of multiple lesions (3 and 6) were obtained. The patients ranged in age from 2 to 40 years, with a mean ± SD of 12.4 ± 11.8 years.

3.2. Measurement of surface area, area, and volume in 2D photographs or 3D images traced by dermatologists and QuantifiCare operators

The measurement results for surface area, area, and volume obtained from 2D photographs and 3D images (traced by all raters independently) are shown in Supplementary Tables S2-S4 (*https://www.ddtjournal.com/action/getSupplementalData.php?ID=287*), respectively. Among patients A-I, A-E had VM, F-H had LM, and I had TA. Patient E with VM had 3 lesions (E1-3), while patient H with LM had 6 lesions (H1-6). When measuring case E1, the traced area was too small to evaluate in 2nd trace of dermatologist rater 2 and 2nd trace of operator rater 4, resulting in a measurement of '0.0' for the surface area and volume. Also, for example, case G1 showed considerable variation among dermatologist rater 1-3 (Figure 3a). Case I1 of dermatologist rater 1 also showed nearly 2-fold divergence between 1st and 2nd traces (Figure 3b). Furthermore, case B1 showed a 10-fold difference between dermatologists and operators (Figure 3c), with the operators detecting and tracing the raised portion of the lesion in the 3D image (Figure 3d).

A summary of the surface area, area and volume measurements by the dermatologists is shown in Table 1. The mean values of surface area for rater 2 were the lowest in both 1st (169.59 mm^2) and 2nd trace (222.86 mm^2), whereas those for rater 3 were the highest in both 1st (446.57 mm^2) and 2nd trace (466.23 mm^2). The mean area measurements also varied several fold among the raters, and the mean volume differed by nearly tenfold among them. However, median and geometric mean values of surface area were relatively similar among the three dermatologists. This discrepancy may be due to the influence of outliers or distortions in the data distribution.

By contrast, the surface area, area, and volume measurements by the QuantifiCare operators were generally consistent with each other (Table 1).

3.3. Analysis of ICC in the evaluation of surface area determined by dermatologists and operators

Partial intra-rater reliability of surface area measured by

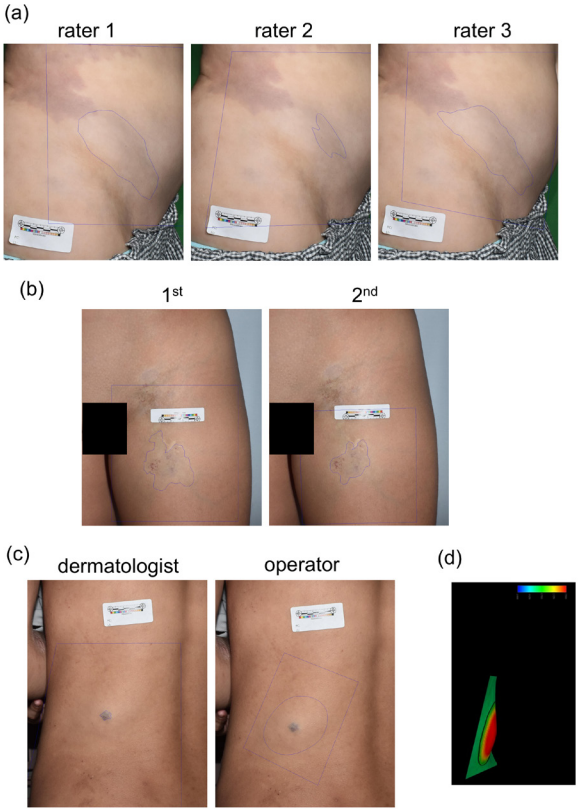

Figure 3. Examples of cases with variable tracing results. (a) Trace of 2D photographs for case G1 by dermatologist rater 1-3. **(b)** Trace of 2D photographs for case I1 by dermatologist rater 1 twice (1st and 2nd). **(c)** Trace of 2D photographs for case B1 by dermatologist rater 3 and QuantifiCare operator 6. **(d)** 3D images color-coded by depth and traced by QuantifiCare operator 6.

Table 1. Summary of measurements of surface area, area and volume evaluated by all raters

		dermatologists		
		Rater 1	Rater 2	Rater 3
Surface Area (mm^2)				
1st	Number of lesions	16	16	16
	Mean (SD)	398.02 (778.84)	169.59 (216.61)	446.57 (801.38)
	Median [Min, Max]	86.22 [17.2, 2810.7]	70.91 [14.4, 691.8]	68.82 [15.2, 2762.7]
	Geo-Mean (Geo-SD)	106.37 (4.82)	81.28 (3.56)	114.10 (5.35)
	CV%	195.7	127.7	179.5
2nd	Number of lesions	16	16	16
	Mean (SD)	339.82 (585.78)	222.86 (394.94)	466.23 (841.60)
	Median [Min, Max]	71.58 [11.3, 1718.3]	78.29 [0.0, 1582.7]	77.47 [15.0, 2966.6]
	Geo-Mean (Geo-SD)	93.17 (4.99)	98.85 (3.72)	121.34 (5.33)
	CV%	172.4	177.2	180.5
Area (mm^2)				
1st	Number of lesions	16	16	16
	Mean (SD)	357.49 (635.95)	161.11 (198.17)	415.87 (678.56)
	Median [Min, Max]	69.20 [19.1, 2041.0]	57.20 [19.0, 690.9]	80.70 [15.9, 2001.4]
	Geo-Mean (Geo-SD)	111.94 (4.40)	84.02 (3.22)	118.75 (5.08)
	CV%	177.9	123.0	163.2
2nd	Number of lesions	16	16	16
	Mean (SD)	324.92 (540.16)	194.44 (275.66)	429.03 (699.15)
	Median [Min, Max]	62.30 [15.8, 1581.0]	56.65 [18.7, 1016.7]	82.45 [15.3, 2160.6]
	Geo-Mean (Geo-SD)	99.05 (4.63)	89.25 (3.52)	126.30 (5.01)
	CV%	166.2	141.8	163.0
Volume (mm^3)				
1st	Number of lesions	16	16	16
	Mean (SD)	575.16 (1453.40)	65.76 (105.76)	578.58 (1406.47)
	Median [Min, Max]	11.87 [0.3, 4810.1]	11.11 [0.4, 363.2]	8.17 [0.4, 5224.2]
	Geo-Mean (Geo-SD)	17.25 (18.85)	11.16 (9.91)	21.40 (21.45)
	CV%	252.7	160.8	243.1
2nd	Number of lesions	16	16	16
	Mean (SD)	456.38 (1111.10)	79.98 (143.42)	531.33 (1165.03)
	Median [Min, Max]	10.06 [0.1, 4004.1]	14.10 [0.0, 468.3]	13.48 [0.2, 4142.7]
	Geo-Mean (Geo-SD)	14.62 (22.63)	15.74 (9.56)	24.54 (21.92)
	CV%	243.5	179.3	219.3
		QuantifiCare operators		
		Rater 4	Rater 5	Rater 6
Surface Area (mm^2)				
1st	Number of lesions	16	16	16
	Mean (SD)	530.57 (823.22)	572.91 (890.87)	536.50 (773.38)
	Median [Min, Max]	115.17 [15.6, 3008.7]	116.99 [20.5, 3133.1]	108.07 [16.9, 2365.3]
	Geo-Mean (Geo-SD)	144.92 (5.99)	164.56 (5.57)	148.07 (5.98)
	CV%	155.2	155.5	144.2
2nd	Number of lesions	16	16	16
	Mean (SD)	468.56 (781.52)	571.37 (872.27)	604.49 (910.73)
	Median [Min, Max]	88.87 [0.0, 2959.3]	114.88 [20.1, 3020.5]	82.28 [17.9, 2875.4]
	Geo-Mean (Geo-SD)	142.83 (5.80)	160.21 (5.73)	154.47 (6.15)
	CV%	166.8	152.7	150.7
Area (mm^2)				
1st	Number of lesions	16	16	16
	Mean (SD)	481.06 (710.90)	545.62 (796.77)	522.15 (725.48)
	Median [Min, Max]	103.75 [16.8, 2452.6]	122.70 [21.8, 2696.4]	98.40 [17.6, 2118.4]
	Geo-Mean (Geo-SD)	140.00 (5.68)	171.38 (5.27)	155.19 (5.67)
	CV%	147.8	146.0	138.9
2nd	Number of lesions	16	16	16
	Mean (SD)	424.87 (660.56)	549.37 (796.72)	575.14 (832.76)
	Median [Min, Max]	94.50 [15.2, 2346.1]	112.15 [21.9, 2633.2]	108.20 [18.1, 2478.5]
	Geo-Mean (Geo-SD)	122.44 (5.63)	169.14 (5.34)	159.81 (5.79)
	CV%	155.5	145.0	144.8

SD, standard deviation; Min, minimum; Max, maximum; Geo-Mean, geometric mean; Geo-SD, geometric standard deviation; CV, coefficient of variation.

Table 1. Summary of measurements of surface area, area and volume evaluated by all raters (continued)

		QuantifiCare operators		
		Rater 4	Rater 5	Rater 6
Volume (mm^3)				
1st	Number of lesions	16	16	16
	Mean (SD)	1008.51 (2028.10)	1199.63 (2366.48)	1135.07 (2099.54)
	Median [Min, Max]	30.28 [0.3, 7214.0]	37.45 [0.4, 7974.8]	31.22 [0.3, 6445.4]
	Geo-Mean (Geo-SD)	34.17 (31.40)	43.81 (26.70)	37.72 (29.97)
	CV%	201.1	197.3	185.0
2nd	Number of lesions	16	16	16
	Mean (SD)	825.75 (1815.78)	1239.12 (2454.64)	1374.67 (2593.08)
	Median [Min, Max]	20.78 [0.0, 6591.1]	33.59 [0.4, 8193.4]	15.99 [0.3, 7863.7]
	Geo-Mean (Geo-SD)	33.51 (29.59)	42.98 (28.27)	39.62 (31.34)
	CV%	219.9	198.1	188.6

SD, standard deviation; Min, minimum; Max, maximum; Geo-Mean, geometric mean; Geo-SD, geometric standard deviation; CV, coefficient of variation.

Table 2. Intra-rater and inter-rater reliability of surface area measurements by dermatologists

Partial Intra-rater reliability					
	Rater	Estimate of ICC (1, 1)	95% CI		SE
			Lower	Upper	
surface area (mm^2)	Rater 1	0.894	0.730	0.961	224.101
	Rater 2	0.702	0.344	0.884	173.625
	Rater 3	0.991	0.974	0.997	79.313
Partial Inter-rater reliability					
Parameter	Measurement	Estimate of ICC (2, 1)	95% CI		SE
			Lower	Upper	
surface area (mm^2)	1st	0.669	0.414	0.854	383.949
	2nd	0.804	0.609	0.920	284.484
Overall Inter-rater and Intra-rater reliability					
Parameter		Estimate	95% CI		SE
			Lower	Upper	
surface area (mm^2)	ρ_{inter}	0.746	0.578	1.000	330.137
	ρ_{intra}	0.933	0.822	1.000	169.958

ICC, intraclass correlation coefficients; CI, confidence interval; SE, standard error.

the dermatologists was 0.894 for rater 1, 0.702 for rater 2, and 0.991 for rater 3 (Table 2), while inter-rater reliability was 0.669 for the 1st trace and 0.804 for the 2nd trace (Table 2). In the analysis of overall inter-rater and intra-rater reliability with repeated measurements of surface area, the correlation coefficient between raters was quite high at 0.746 (interpretation: Substantial agreement), and the correlation coefficient within raters was very high at 0.933 (Almost perfect agreement) (Table 2).

On the other hand, partial intra-rater reliability of surface area measured by the QuantifiCare operators was 0.985 for rater 4, 0.995 for rater 5, and 0.961 for rater 6 (Table 3), while inter-rater reliability was 0.938 for the 1st trace and 0.954 for the 2nd trace (Table 3). In the analysis of overall inter-rater and intra-rater reliability with repeated measurements of surface area, the correlation coefficient between raters was quite high at 0.946 (Almost perfect agreement), and the correlation coefficient within raters was very high at 0.981 (Almost perfect agreement) (Table 3). Thus, all values obtained by the QuantifiCare operators were greater than those obtained by the dermatologists.

3.4. Analysis of ICC in the evaluation of area

Partial intra-rater reliability and inter-rater reliability of area evaluated by dermatologists were similar to those for surface area (Supplementary Table S5-S6, *https://www.ddtjournal.com/action/getSupplementalData.php?ID=287*). In the analysis of overall inter-rater and intra-rater reliability with repeated measurements of area, the correlation coefficient between raters was quite high

Table 3. Intra-rater and inter-rater reliability of surface area measurements by QuantifiCare operators

Partial Intra-rater reliability					
	Rater	Estimate of ICC (1, 1)	95% CI		SE
			Lower	Upper	
surface area (mm^2)	Rater 4	0.985	0.960	0.995	97.380
	Rater 5	0.995	0.987	0.998	60.444
	Rater 6	0.961	0.895	0.986	166.702
Partial Inter-rater reliability					
Parameter	Measurement	Estimate of ICC (2, 1)	95% CI		SE
			Lower	Upper	
surface area (mm^2)	1^{st}	0.938	0.865	0.976	205.830
	2^{nd}	0.954	0.896	0.982	184.411
Overall Inter-rater and Intra-rater reliability					
Parameter		Estimate	95% CI		SE
			Lower	Upper	
surface area (mm^2)	ρ_{inter}	0.946	0.902	1.000	195.200
	ρ_{intra}	0.981	0.954	1.000	116.799

ICC, intraclass correlation coefficients; CI, confidence interval; SE, standard error.

at 0.723 (Substantial agreement), and the correlation coefficient within raters was very high at 0.952 (Almost perfect agreement) (Supplementary Table S7, *https://www.ddtjournal.com/action/getSupplementalData.php?ID=287*).

Partial intra-rater reliability, inter-rater reliability, and overall inter-rater and intra-rater reliability with repeated measurements of area evaluated by QuantifiCare operators were also similar to those of surface area (Supplementary Table S8-S10, *https://www.ddtjournal.com/action/getSupplementalData.php?ID=287*), and all values obtained by the QuantifiCare operators were higher than those by dermatologists.

3.5. Analysis of ICC in the evaluation of volume

Similar to the evaluation of surface area and area, all values evaluated by QuantifiCare operators were greater than those by dermatologists (Supplementary Table S11-S16, *https://www.ddtjournal.com/action/getSupplementalData.php?ID=287*).

4. Discussion

We previously conducted a randomized phase II clinical trial of topical sirolimus therapy for cutaneous lesions of VM, LM, TA, and kaposiform hemangioendothelioma (*11*). The primary endpoint of the study was the overall improvement score in the target lesion (size and coloration) assessed according to photographs by the independent review committee at Week 12. However, there was no statistically significant difference in the mean improvement score in the 0.2% sirolimus gel group or the 0.4% group compared with the placebo group. Thus, we could not prove the efficacy of topical sirolimus for cutaneous VA in the protocol.

On the other hand, among secondary endpoints, the improvement in target lesion size at Week 12 as assessed by the independent review committee was significantly higher in the 0.4% sirolimus gel group than in the placebo group. Also, a mixed-effects model for repeated measures analysis showed a significant difference in the major diameter of lesion between the placebo and 0.4% groups at Week 16. We therefore hypothesized that the focus should be on size rather than coloration for the evaluation of efficacy against skin lesions of VA. Accordingly, the method of measuring the area of the target lesion was changed from calculating the area from the multiplication of major and minor diameters to tracing the contour of the lesion, which resulted in remarkable differences in some cases. As a result, the post-hoc analysis indicated that the percentage of patients with ≥ 20% reduction in the re-measured lesion area was significantly higher in the sirolimus groups at Week 8 and 12. Thus, the efficacy of sirolimus gel by the post-hoc analysis could be detected earlier than with the secondary endpoints (Week 12 and 16). However, this post-hoc re-measurement did not capture elevation of the lesions.

Based on these results of the clinical trial, we needed to find a better method to measure lesion size for the next phase III trial. The procedure in the present study involved tracing lesion contours on photographs of target diseases of the phase II clinical trial captured with a 3D camera, followed by 3D processing of the images to measure lesion area, surface area, and volume (*6-8*).

Comparisons were made between measurements based on 2D photographs traced by dermatologists and those based on 3D images traced by QuantifiCare operators. Both methods yielded high inter- and intra-rater reliability; however, the inter- and intra-rater reliability for measurements by the QuantifiCare operators was even higher.

One possible explanation is that tracing a 2D photograph relies only on planar information, whereas tracing a 3D image was based on three-dimensional information. This additional depth information makes it easier to discern a lesion's uneven surface and identify its true contour more accurately. The dermatologist raters gave feedback such as: "Several lesions have abruptly broken edges and others gradually become normal with a gradation. It is very difficult to determine the contour of such lesions with the naked eye." and "Some lesions have indistinct borders, resulting in variations between the 1st and 2nd tracings.".

Despite clinical needs, the specific problems of applying 3D imaging technology to measurement of VA included a variety of clinical presentations of lesions ranging from superficial macules to deep masses with unevenness, blurred boundaries, color similarity to surrounding tissues, and irregular/lobulated morphology. In the present study, we attempted to address these challenges by using a dedicated 3D camera system that generates color-coded depth maps, by standardizing patient positioning and camera distance, and by having operators trace the lesions on 3D images.

Given the higher inter- and intra-rater reliability of the 3D measurement method in this study, if a 3D-based quantitative endpoint such as surface area or volume had been adopted in the phase II trial, the drug efficacy might have been detected with greater sensitivity or at an earlier time point.

5. Limitations

This study has several limitations. First, in the case of E1, surface area and volume evaluated by two raters were "0.0" but area was "20.9" or "15.3". In the 3D measurement of surface area and volume, a dense mesh (or grid composed of several triangles) is applied to the reconstructed 3D surface. Then, the system calculates the surface area and volume based on how many triangles are fully covered by the lesion. However, there is a detectability limit when applying the mesh. The contours traced by the two raters did not cover a sufficient number of mesh triangles, resulting in those triangles being excluded and the measurement becoming 0.0. This discrepancy indicates that lesions must be above a certain size (to stay above the detection limit), when using 3D measurement in future practice through standardized operating procedures.

Second, there were cases where the surface area was smaller than the area (*e.g.* case A1 and H6). This was also due to the calculation method, and it is known that large variations can occur depending on (i) the photography angle, (ii) the flatness of the lesion and its surrounding area, and (iii) whether the calibration sticker is placed in exactly the same plane as the lesion. These imaging conditions should be carefully considered in future studies to achieve more accurate measurements.

Next, it is possible that the 3D method is more advantageous for prominently raised VM lesions, while offering less benefit for flat LM lesions. According to our subgroup analysis, no correlation was observed between the 3D measurement reliability and disease type, size, or unevenness of lesions (data not shown). However, caution is required in interpreting this subgroup analysis due to the limited sample size.

Lastly, the cost-effectiveness of this 3D measurement method also needs to be considered. From the viewpoint of clinical translation feasibility, in future clinical trials the contract research organization will provide the participating medical institutions with cameras, a subscription to the 3D visualization platform, and initial training for physicians and staff. According to the manufacturer, the initial cost includes a dedicated 3D camera system and a software license (with an annual software fee and warranty), which is higher than the cost of a standard digital camera. Nonetheless, the improved reproducibility and objectivity of lesion-size measurements may justify the upfront cost by enabling more efficient clinical trials and more reliable assessment of treatment response.

6. Conclusion

We compared size measurements of cutaneous VA lesions obtained from 2D photographs traced by three dermatologists with those obtained from 3D images traced by three company operators, and assessed inter- and intra-rater reliability. Both 2D and 3D methods demonstrated high inter- and intra-rater reliability; however, better reliability was observed in the measurements obtained by company operators using 3D imaging. The findings indicate that 3D surface imaging provides a more consistent and objective evaluation of lesion size than 2D photography, and support the potential application of this method in clinical practice and clinical trials for VA.

Acknowledgements

We would like to thank all investigators and secretaries of participating institutes of the present study. We acknowledge proofreading and editing by Benjamin Phillis at the Clinical Study Support Center at Wakayama Medical University, and the assistance in analyzing the data by QuantifiCare and Nobelpharma.

Funding: This study was funded by the research budget

of Wakayama Medical University. However, labor costs and subcontracting expenses related to statistical analysis and measurement were funded by Nobelpharma Co., Ltd.

Conflict of Interest: This study was funded by Nobelpharma. Masatoshi Jinnin received consultancy fee from Nobelpharma.

References

1. International Society for the Study of Vascular Anomalies. ISSVA classification of vascular anomalies[Internet]. 2025[cited 2025 May 28]. Available from: *https://www.issva.org/classification* (accessed May 28, 2025).
2. Kunimoto K, Yamamoto Y, Jinnin M. ISSVA Classification of Vascular Anomalies and Molecular Biology. Int J Mol Sci. 2022; 23:2358.
3. Ozeki M, Nozawa A, Yasue S, Endo S, Asada R, Hashimoto H, Fukao T. The impact of sirolimus therapy on lesion size, clinical symptoms, and quality of life of patients with lymphatic anomalies. Orphanet J Rare Dis. 2019; 14:141.
4. Ozeki M, Endo S, Yasue S *et al*. Sirolimus treatment for intractable lymphatic anomalies: an open-label, single-arm, multicenter, prospective trial. Front Med (Lausanne). 2024; 11:1335469.
5. Singh S, Bradford D, Li X *et al*. FDA Approval Summary: Alpelisib for PIK3CA-Related Overgrowth Spectrum. Clin Cancer Res. 2024; 30:23-28.
6. Chaby G, Lok C, Thirion JP, Lucien A, Senet P. Three-dimensional digital imaging is as accurate and reliable to measure leg ulcer area as transparent tracing with digital planimetry. J Vasc Surg Venous Lymphat Disord. 2017; 5:837-843.
7. Bharadia SK, Gabriel V. Comparison of Clinical Estimation and Stereophotogrammic Instrumented Imaging of Burn Scar Height and Volume. Eur Burn J. 2024; 5:38-48.
8. Bonan P, Verdelli A. Combined microwaves and fractional microablative CO2 laser treatment for postpartum abdominal laxity. J Cosmet Dermatol. 2021; 20:124-131.
9. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979; 86:420-428.
10. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas. 1973; 33:613-619.
11. Jinnin M, Shimokawa T, Kishi A *et al*. Topical sirolimus therapy for cutaneous vascular anomalies: A Randomized Phase II Clinical Trial. J Dermatol. 2025; 52:872-887.

Received September 27, 2025; Revised January 19, 2026; Accepted February 13, 2026.

**Address correspondence to:*
Kayo Kunimoto, Department of Dermatology, Wakayama Medical University, 811-1 Kimiidera, Wakayama 641-0012 Japan.
E-mail: k-jigen@wakayama-med.ac.jp

Released online in J-STAGE as advance publication February 20, 2026.