Volume Measures Using a Digital Image Analysis System are Reliable in Diabetic Foot Ulcers

Sue E. Gardner, PhD, RN[1], **Rita A. Frantz, PhD, RN, FAAN**[1], **Stephen L. Hillis, PhD**[2], **Tom J. Blodgett, MSN, RN, PhD(c)**[1], **Lorraine M. Femino, BSN**[1], and **Shannon M. Lehman, BBA, BSW**[1]

[1]University of Iowa College of Nursing, Iowa City, IA

[2]Comprehensive Access and Delivery Research and Evaluation (CADRE) Center, Iowa City VA Medical Center

Abstract

Reliable measures of wound size are critical to wound healing research and clinical management. Measurement of full-thickness wounds is increasingly being done using digital images and photogrammetric software, such as VeVMD (Vista Medical, Winnipeg, Manitoba, Canada), to estimate wound volume. The reliability of VeVMD in determining wound volume is unknown. The present study sought to examine the reliability of wound volume measurements obtained using VeVMD.

Methods—A cross-sectional study of adults with full-thickness, neuropathic, diabetic foot ulcers (DFU) at 2 sites in the US Midwest was undertaken. Ulcer images were obtained, stored, and used to obtain measures of wound volume using VeVMD. Four raters independently completed wound measures, and then repeated these measures 2 weeks after the first measurement. Raters were blinded to the comparison measurements. Inter- and intra-rater correlations were computed.

Results—Thirty-three enrolled subjects with 33 DFU were included in the analyses. Inter-rater reliability was 0.745 and intra-rater reliability was 0.868. Four ulcers showed noticeably less agreement between raters; these ulcers had small, but deeply recessed areas, resulting in differences in defining the wound margin. When these 4 ulcers were removed, inter- and intra-rater reliabilities were excellent (0.970 and 0.981, respectively).

Conclusion—Reliabilities of volume measurements obtained with VeVMD were acceptable in DFU, even when raters had different definitions of the ulcer margin or changed their definition from time to time. However, conclusions cannot be drawn regarding the performance of VeVMD in other wound types.

The ability to obtain reliable measures of wound size is a critical component of wound healing research and clinical wound management. Compared to measures of surface area and circumference, measures of volume provide a more complete reflection of wound healing progress in full-thickness wounds.[1] Although several techniques are available for measuring volume, such as stereophotography,[2] Kundin device,[3-5] and alginate molds,[6] measurement of full-thickness wounds healing by secondary attention is increasingly being done using digital images of the wound and photogrammetric software that estimates wound

volume since it is less invasive and complicated than other techniques.[7] VeVMD (Vista Medical, Winnipeg, Manitoba, Canada) is a proprietary, computerized technique that is relatively easy to use, and does not require direct contact with the wound. However, the reliability of VeVMD in determining wound volume remains unknown.

The reliability of a measurement technique concerns the extent to which it yields consistent results with repeated measurement.[8] Dimensions of reliability pertinent to computerized techniques to obtain measures of wound size, such as VeVMD, include inter-rater reliability and intra-rater reliability. Inter-rater reliability is important because clinical and research environments often dictate that different persons (or raters) perform wound measures over time. It is not realistic that a single rater will always be available. Therefore, it is critical to establish that an acceptable level of agreement exists among raters. Despite the fact that multiple raters are often involved in measures of wound size, it is also important to establish intra-rater reliability, or the consistency of repeated measures by the same rater.

Several factors may contribute to unreliable determinations of wound volume using computerized analyses of digital images, and VeVMD specifically. One contributor to variable measures is the dependence of VeVMD on the rater to define the outline of the wound margin and depth. This process will differ between raters and within the same rater, and may be influenced by lighting, glare, and image focus. Another factor that can lead to unreliable volume measures is variation in patient and/or wound position that can occur between measurements by the same rater or between measurements of various raters. Patient/wound position may alter the wound margin shape and wound bed depth, both of which are critical to the VeVMD technique. All of these factors reduce the reliability of volume measures. Therefore, reliability needs to be examined to evaluate the dependability of VeVMD in determining wound volume.

The purpose of this study was to examine the reliability of wound volume measures obtained using VeVMD in a sample of plantar-surface diabetic foot ulcers (DFU). Plantar surface DFU are typically circular or elliptical and rarely present with irregular borders. Nevertheless, the careful monitoring of wound healing progress, achieved by reliable wound measures, is imperative in order to prevent complications, such as amputation. The specific aims of this study were to:

1. Identify the inter-rater reliability of DFU volume measurement using VeVMD.

2. Identify the intra-rater reliability of DFU volume measurement using VeVMD.

The current study was completed as part of a larger prospective study to examine bioburden predictors of DFU outcomes (NINR R01 NR009448 PI. Sue Gardner). A cross-sectional research design was used.

Methods

Setting and sample

Subjects in the primary study were recruited from a large academic medical center and a Department of Veterans' Affairs Medical Center, both in the US Midwest. Adults with diabetes who were not taking long-term antibiotics and had a stable, nonvascular neuropathic ulcer on the plantar surface of the foot were recruited for participation in the study. Subjects with evidence of osteomyelitis were excluded. Subjects with full-thickness DFU (ie, extends through the dermal layer) were included in the primary study. Subjects from the primary study with superficial ulcers were excluded, due to the miniscule volume of these ulcers. Approval of the study was obtained from the institutional review board at

both institutions. All subjects provided written, informed consent prior to study participation.

Wound volume measures

The primary study variable was wound volume, measured using the VeVMD software system. This software was downloaded on a Dell Latitude D630 laptop computer (Dell, Round Rock, TX). Digital images of study ulcers were obtained using a Canon PowerShot SX100 camera (Canon USA, Lake Success, NY) at 0.3 megapixel resolution using the "macro" function to enhance image clarity. The camera lens was zoomed in so that the ulcer field occupied the entire viewfinder window. The flash and focus features were set to "automatic." Digital images contained the ulcer, a 3 cm^2 orientation card, the subject ID number, and a single-point wound-depth indicator (ie, a marked cotton-tipped applicator placed at the deepest point of the wound base). At least 6 images were obtained and uploaded to the study computer and into the VeVMD software program. The software oriented each image to scale and to the plane of the ulcer using the 3 cm^2 orientation card. The image with the highest degree of orientation accuracy was selected for all subsequent wound measurements.

Using this image, the ulcer outline was traced using the VeVMD "Shrink-Fit Tool." This tool utilized a series of tracer points, connected by line segments, to outline the edge of the ulcer margin. Depth was measured using the VeVMD "Line Tool," which fits a straight line along the wound depth indicator. The "Zoom" feature in VeVMD was utilized, when necessary, to visualize the edges of the ulcer margin more clearly during measurement. VeVMD estimates volume using the mathematical formula for an elliptical spheroid, which is a function of wound depth and circumference. Wound volume measurements, as well as circumference, surface area, and depth, were recorded on case report forms.

Data collection procedures

Wound images were obtained, as described above, at baseline and every 2 weeks until: 1) the ulcer healed; 2) self or study withdrawal; or 3) 6 months of follow-up had been completed and stored on the study laptop. For this study, only the wound image from baseline was used. The "raters" for this study were 4 different research team members who were trained in: 1) wound assessment, including defining wound margins; 2) obtaining digital images; and 3) completing ulcer measurements using VeVMD software. Detailed study protocols for completing wound size measurements were available for consultation at all times.

Each rater independently completed measures of wound size from the baseline image (Time 1). The order in which each of the raters measured the wound images was generated using a random sequence table. Two weeks after the initial measurement, each rater repeated measurements (Time 2) of the wound images using a different randomly generated order than was used the first time. To minimize bias, each rater deleted his or her ulcer outline and depth line from the selected image on VeVMD before the next rater measured the ulcer.

Analyses of inter- and intra-rater reliability

The units of analysis for the inter- and intra-rater reliability analyses were volume measurements at Time 1 and Time 2. Inter-rater analyses were performed using Time 1 volume measurements. Inter-rater reliability was defined as the correlation between volume measurements from 2 different raters for the same ulcer, at the same time (Time 1). The analyzed data consisted of 4 measurements for each ulcer, corresponding to the 4 different raters. A summary estimate of the inter-rater reliability was computed by averaging the correlations corresponding to each possible pair of raters. An estimated standard error (SE)

for the summary inter-rater reliability estimate was obtained using bootstrap methods[9]; bootstrap samples ($B = 2000$) were drawn, treating ulcers as the sampling units.

Intra-rater reliability was defined as the correlation between Time 1 and Time 2 volume measurements made by the same rater on the same ulcer, using the same wound image for both measurements, which had been obtained at the baseline study visit. A summary estimate of the intra-rater reliability was computed by averaging the 4 within-rater correlations, and standard errors were obtained using bootstrap methods, treating ulcers as the sampling units. Analyses were performed in SAS, Version 9.2 (SAS Institute Inc. Cary, NC).

Results

A total of 34 subjects from the primary study had full-thickness DFU and were eligible for inclusion in this study. Compared to the others, 1 subject had an extremely large ulcer. This subject was excluded from further analyses because the ulcer could influence the results of the study much more than the other ulcers. Of the remaining 33 subjects, the mean age of subjects was 54.7 years (± 11.27). Subjects were predominantly white (90.9%), non-Hispanic (93.9%), and male (90.9%).

Twenty-eight of the 33 subjects (84.8%) had just 1 DFU, and 5 subjects (15.2%) had 2 ulcers each. In these cases, the larger of the 2 ulcers was selected for the study. The 33 study ulcers were located in 1 of 7 regions on the plantar foot surface (Table 1). Baseline measures of ulcer circumference, depth, surface area, and volume are provided in Table 1.

Inter-rater reliability

To investigate inter-rater reliability of volume measures, scatter plots were examined to visually assess reliability. Figure 1 displays a matrix of inter-rater scatter plots of Time 1 volume measures for each pairing of raters. The corresponding pairwise correlations are shown in Table 2. The summary inter-rater reliability estimate, computed as the average of the correlations in Table 2, is 0.745 (bootstrap SE = 0.0968).

Intra-rater reliability

To investigate intra-rater reliability of volume measures, scatter plots were examined to visually assess reliability. Figure 2 displays a matrix of inter-rater scatter plots of Time 1 vs. Time 2 volume measures for each rater. The corresponding pairwise correlations are shown in Table 3. The summary estimate of the intra-rater reliability, computed as the average of the correlations in Table 3, is 0.868 (bootstrap SE = 0.0563).

Analyses excluding outliers

Based on the scatter plots represented in Figure 1 and Figure 2, outliers were investigated further to discern possible reasons for these substantial discrepancies between and within raters. Four study ulcers were consistently contributing to these outlier points for both the inter-and intra-rater plots. Typically, these ulcers had a deep recession in a small area, while the remainder was relatively superficial in comparison to the recession (Figure 3).

In comparing ulcer measurements between raters or within raters, it was discovered that the discrepancy was due to differences in defining the ulcer margin. Some raters defined the margin as the small, deep recession only, while others defined it as inclusive of all tissue loss. Additionally, some raters changed the definition of the ulcer margin from their measurements at Time 1 to Time 2.

In order to determine the influence that this common error had on the reliability estimates, the analyses were repeated, excluding the 4 ulcers mentioned above (outliers). The corresponding inter-rater pairwise correlations from Time 1 are shown in Table 4. The summary inter-rater reliability estimate, computed as the average of the correlations in Table 4, is 0.970 (bootstrap SE = 0.0151).

The corresponding intra-rater pair-wise correlations are shown in Table 5. The summary estimate of the intra-rater reliability, computed as the average of the correlations in Table 3, is 0.981 (bootstrap SE = 0.00782).

Discussion

The inter-rater reliability of wound surface area measurements obtained with VeVMD was reported to be between 0.94 and 0.99, and the intra-rater reliability using VeVMD was 0.99.[7] This was the first study to examine the reliability of volume measures using this technique. Based on the findings of inter- and intra-rater analyses that included all study ulcers, the reliability of volume measures obtained with the VeVMD have fair inter-rater reliability (ie, 0.745) and high intra-rater reliability (ie, 0.868).These rates were achieved in this study despite the inclusion of 4 different raters, which is consistent with the number of different team members often involved with the measurement of wound size in both practice and research environments. However, the raters in this study were carefully trained and had the benefit of detailed protocols for wound measurement. Despite this fact, it was found that the most significant form of error related to doing volume measures was defining the ulcer margin, not the ability to use the software to trace the margin. When the ulcers contributing to this type of error were removed from the analysis, both inter- and intra-rater reliability were excellent (0.97 and 0.98, respectively). This finding highlights the need for ongoing review of the definition of "wound margin" and detailed protocols in order to ensure a high degree of measurement reliability when using measurement systems that rely on wound images.

The sample of wounds in this study was limited to plantar neuropathic DFU, which are more uniform in size and shape than other types of wounds. Variations in the regularity and shape of wound margins, depth of the wound base, and wound curvature, can influence the reliability of wound measurement techniques.[10] For example, volume measures of pressure ulcers, which can be quite large, occur on a variety of body surfaces, and can wrap around body surfaces, may not demonstrate the same degree of reliability found for volume measures of DFU. The reliability of volume measures needs to be established for a variety of wound types in order to guide appropriate selection of techniques for a specific clinical or research purpose.

Conclusion

Although reliable measures of wound size are critical to clinical practice and research, the accuracy of these measures is equally, if not more, important. Although this study confirmed that volume measures based on VeVMD are reliable in DFU, reliability does not mean that they are accurate. The accuracy of volume measures based on VeVMD has not been examined in any wound type, including DFU. To complement this study, further study is needed to examine the accuracy of these measures in DFU.

Acknowledgments

Resources (NCRR), a part of the National Institutes of Health (NIH). The views expressed are those of the authors and do not necessarily represent views of the CTSA or NINR.

References

1. Wysocki AB. Wound measurement. Int J Dermatol. 1996; 35(2):82–91. [PubMed: 8850032]

2. Frantz RA, Johnson DA. Stereophotography and computerized image analysis: A three-dimensional method of measuring wound healing. WOUNDS. 1992; 4(2):58–64.

3. Kundin JI. Designing and developing a new measuring instrument. Perioper Nurs Q. 1985; 1(4):40–45. [PubMed: 3852330]

4. Langemo DK, Melland H, Olson B, et al. Comparison of 2 wound volume measurement methods. Adv Skin Wound Care. 2001; 14(4):190–196. [PubMed: 11902344]

5. Johnson M, Miller R. Measuring healing in leg ulcers: practice considerations. Appl Nurs Res. 1996; 9(4):204–208. [PubMed: 8961578]

6. Stotts NA, Salazar MJ, Wipke-Tevis DD, McAdoo E. Accuracy of alginate molds for measuring wound volumes when prepared and stored under varying conditions. WOUNDS. 1996; 8(5):158–164.

7. Thawer HA, Houghton PE, Woodbury MG, Keast D, Campbell K. A comparison of computer-assisted and manual wound size measurement. Ostomy Wound Manage. 2002; 48(10):46–53. [PubMed: 12378003]

8. Strickland, OL.; Lenz, ER.; Waltz, CF. Measurement in Nursing and Health Research. Fourth. New York: Springer Publishing Company; 2010.

9. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. London: Chapmann & Hall; 1993.

10. Little C, McDonald J, Jenkins MG, McCarron P. An overview of techniques used to measure wound area and volume. J Wound Care. 2009; 18(6):250–253. [PubMed: 19661849]

Keypoints

- The purpose of this study was to examine the reliability of wound volume measures obtained using photogrammetric software (VeVMD) in a sample of plantar-surface diabetic foot ulcers (DFU).

- Subjects with full-thickness DFU (ie, extends through the dermal layer) were included in the primary study.

- Each rater independently completed measures of wound size from the baseline image (Time 1). Two weeks after the initial measurement, each rater repeated measurements (Time 2) of the wound images using a different randomly generated order than was used the first time.

- In comparing ulcer measurements between raters or within raters, it was discovered that the discrepancy was due to differences in defining the ulcer margin. Some raters defined the margin as the small, deep recession only, while others defined it as inclusive of all tissue loss.

- Based on the findings of inter- and intra-rater analyses that included all study ulcers, the volume measures obtained with the VeVMD had fair inter-rater reliability (ie, 0.745) and high intra-rater reliability (ie, 0.868).

- The sample of wounds in this study was limited to plantar neuropathic DFU, which are more uniform in size and shape than other types of wounds. Variations in the regularity and shape of wound margins, depth of the wound base, and wound curvature, can influence the reliability of wound measurement techniques.[10]
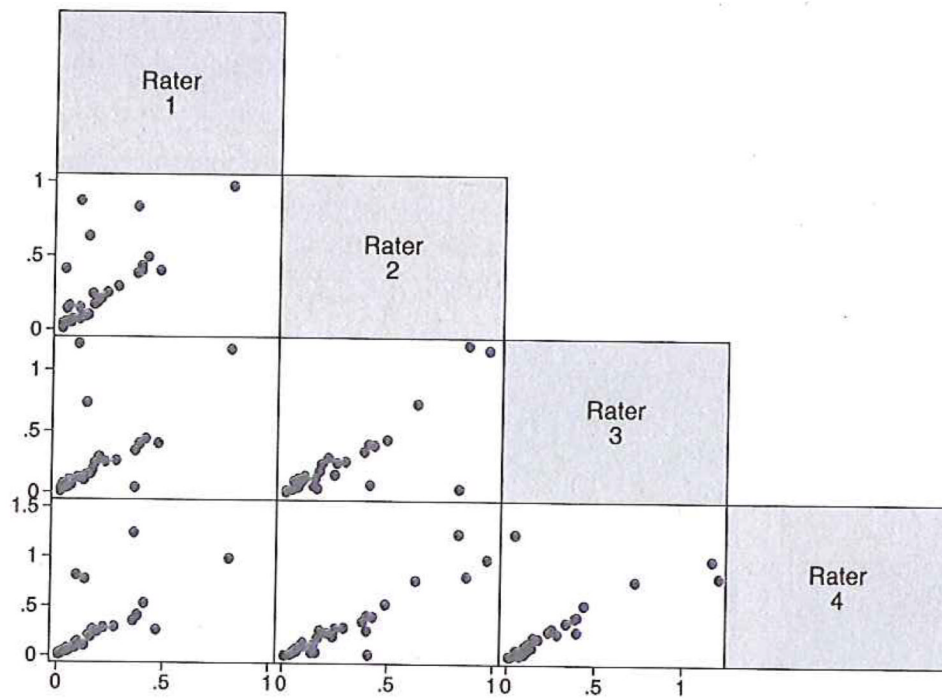
Figure 1.
Inter-rater scatter plot for Time 1 volume measures (n = 33). Each point represents 2 volume measurements at Time 1 from the specified pair of raters.
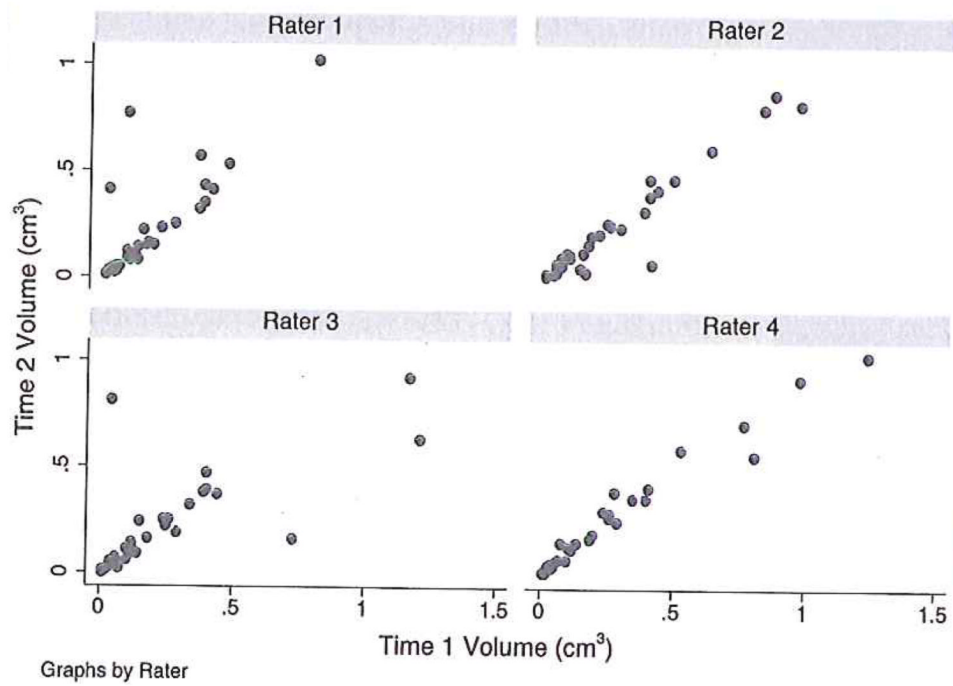
Figure 2.
Intra-rater scatter plot of Time 1 and Time 2 volume measurements (n = 33). Each point represents the volume measures at Time 1 and Time 2 for the same ulcer by the specified rater.

Figure 3.
Superficial ulcer with non-uniform depth in wound base.

Table 1

Baseline characteristics of the study ulcers.

Ulcer Characteristic	N = 33
Location, frequency (%)	
Hallux	10 (30.3%)
Central forefoot	9 (27.3%)
Medial midfoot	4 (12.1%)
Hindfoot	4 (12.1%)
Medial midfoot	3 (9.1%)
Lateral forefoot	2 (6.1%)
Lateral midfoot	1 (3.0%)
Mean wound size (± SD)	
Circumference (cm)	4.6 (± 2.37)
Surface area (cm^2)	1.5 (± 1.67)
Depth (cm)	0.44 (± 0.23)
Volume (cm^3)	0.20 (±0.22)

Table 2

Inter-rater correlations for Time 1 volume measures (n = 33).

	Rater 1	**Rater 2**	**Rater 3**
Rater 2	0.71		
Rater 3	0.62	0.82	
Rater 4	0.69	0.93	0.71

Table 3

Intra-rater correlations between Time 1 and Time 2 volume measures (n = 33).

Rater	Correlation
1	0.80
2	0.96
3	0.73
4	0.98

Table 4

Inter-rater correlations for Time 1 volume measures, excluding outliers (n = 29).

	Rater 1	Rater 2	Rater 3
Rater 2	0.97		
Rater 3	0.97	0.97	
Rater 4	0.96	0.98	0.98

Table 5

Intra-rater correlations between Time 1 and Time 2 volume measures, excluding outliers (n = 29).

Rater	Correlation
1	0.98
2	0.98
3	0.98
4	0.99