# Reliability study of a new wound classification system for patients with diabetes

## ClinicalPractice

**Abstract**

**Introduction:** Wound classification systems, specifically for diabetic ulcer patients, have been less explored in reliability studies than other types, such as pressure ulcers. Therefore, this study evaluates the inter-rater reliability of a new wound classification system developed for diabetic foot ulcers in Indonesia.

**Method:** This is a cross-sectional pilot study carried out to determine the inter-rater reliability of the Suriadi, Haryanto, Imran dan Defa (SHID) wound classification system at the Kitamura Wound Clinic in Pontianak city, Indonesia. Fifteen adult outpatients with diabetic ulcers were asked to participate in the evaluation of the wound classification systems and forty wound photographs were used.

**Results:** The result showed that the inter-rater reliability scores of the SHID were in perfect agreement (kappa [κ]=0.81-1.00) as opposed to the Wagner (κ=0.43-0.77) and the Texas (κ=0.34-63) wound classification systems.

**Conclusion:** The SHID wound classification system can help identify severe wounds for treatment, especially among people with diabetic ulcerss.

**Keywords:** wound classification, SHID, inter-rater reliability, diabetic wounds

**Abbreviations:** PEDIS: Perfusion, Extent, Depth, Infection and Sensation, SINBAD: Site, Ischaemia, Neuropathy, Bacterial Infection, Area and Depth, UT: University of Texas, WCN: Wound Care Nurses, WIFI: Wound Infection, Foot Infection

## Introduction

Endemic infections have been an issue Wounds are common complications and major sources of morbidity among patients with diabetes [1], who have a 15-40 times higher risk for lower extremity amputation [2]. The regional incidence of amputation due to wounds among patients with diabetes in Indonesia has been estimated at 41.5% [3]. As such, identifying the wound classification is particularly important in the management of diabetic ulcers in order to determine the progress of wound healing and prevent further complications, such as amputation. Early and accurate assessments of diabetic wounds can help reduce adverse effects and prevent incidents of amputation. Accordingly, several diabetic foot wound classification systems have been developed and studied in an effort to create an effective system for wound management and treatment. Such widely circulated wound classification systems include the Curative Health Care System, a descriptive wound classification system that has six levels and resembles a modified Wagner scale; Diabetic Foot Risk Assessment Tool; Scottish Intercollegiate Guidelines Network/Scottish Care Information-Diabetes Collaboration; Pickwell et al. simplified system; The Size (Area, Depth), Sepsis, Arteriopathy and Denervation system; and Van Acker-Peter classification system modified from the University of Texas system. Nonetheless, no reliability studies have been available for the aforementioned classification systems [4]. Wound, Infection, Foot Infection (WIFI) is another classification system that has been shown to predict multiple pertinent diabetic foot ulcer outcomes [4]. However, only one reliability study demonstrated high levels of inter- and intra-observer reliability, which, unfortunately, did not include all patients with diabetes [5]. Moreover, to utilise the WIFI system, clinicians require vascular equipment for identifying ankle-brachial pressure index and transcutaneous oxygen tension measurements, with such an assessment becoming a waste of time when no or little peripheral vascular disease

**Suriadi***,
**Haryanto,**
**Imran and**
**Defa Arisandi**

Post Graduate Nursing Program, The Institute of Nursing Muhammadiyah Pontianak, Indonesia

*Author for correspondence:

suriadif@yahoo.com.au

is observed, especially in patients with diabetic ulcers [6]. Furthermore, this tool may not be available, especially in countries with limited health care service facilities, such as Indonesia.

The Wagner and University of Texas (UT) systems are two of the most widely used ulcer classification systems [1]. Although the Wagner classification system has been widely used in studies, evidence regarding its reliability has been lacking [4,5]. Another study also reported that the Wagner classification does not adequately address all diabetic foot ulcerations and infections, including vascular problems [6]. The UT system, on the other hand, has been validated to be generally predictive of outcome given that wounds with a greater grade and stage are less likely to heal without revascularisation or amputation. Inter-rater reliability studies have determined that the Wagner (kappa=0.42-0.55) and UT scale (kappa=0.46-0.51) have similar kappa coefficients [7]. Meanwhile, other classification systems, namely the king system; Perfusion, Extent, Depth, Infection and Sensation (PEDIS) scale; and the Site, Ischaemia, Neuropathy, Bacterial Infection, Area and Depth (SINBAD) system, have been available [8]. Accordingly, the PEDIS scale had an inter-rater reliability kappa coefficient of 0.574 [95% CI (confidence interval): 0.522-0.626] [7], while the UT, SINBAD and PEDIS scales exhibited slight to moderate single-observer reliability, with kappa coefficients of 0.53, 0.44 and 0.23-0.42, respectively, except for multiple observers, which presented perfect agreement [8]. The aforementioned study concluded that reliability studies conducted on several wound classification systems have produced unsatisfactory results. One study found that four classification systems, namely the PEDIS, Wagner, SINBAB and UT, had satisfactory inter-rater agreement between three observers calculated using Kendall's Tau coefficient, with the strength of the agreement varying from moderate to almost perfect. The study reported that the Wagner, SINBAD, PEDIS and UT systems had Kendal Tau coefficients of 0.88-0.94, 0.767-0.850, 0.511-0.896 and 0.629-0.942, respectively [9], suggesting that the Wagner system had better reliability compared to the SINBAD, PEDIS and UT systems. However, this study still presents unsatisfactory given the weakness of the Wagner system [4] and reliability testing, especially when using Kendal Tau statistical analysis. Notably, studies have recommended the utilisation of systems with coefficient values of 0.9 and above for medicine [10,11]. A review of the literature concluded that a specific, universally accepted, gold standard classification system has yet to be established.

Consistency remains key in achieving good reliability when using wound assessment tools [12]. Indonesia has several practitioners, such us nurses and medical doctors, who engage in independent practice across the country. Such practitioners provided treatment for chronic ulcers, such as diabetic ulcers, and often make referrals to hospitals when comprehensive treatment is needed. Thus, practitioners need to identify wound classification systems with good reliability early in the treatment process in order to provide proper wound care and prevent complications, such as amputations, especially in patients with diabetes. The current study therefore aimed to evaluate the inter-rater reliability of a new diabetic ulcer classification system called the SHID.

## ■ Development of a new wound classification system

We herein created a new model for determining wound classification called the SHID, which is named after the individuals who created it. The new model was developed based on the authors' clinical observations of patients with diabetic ulcer in Indonesia **(TABLE 1)**. The SHID can be said to be a modification of the UT and Wagner classifications, which both consist of six levels. This model not only describes the levels of layers from skin tissue to the bone but also adds items addressing complications, such us infection, ischaemia and osteomyelitis that accompany skin tissue and bone damage often seen in patients with diabetic ulcers. The first classification characterises the superficial area covering the epidermis layer and/or the dermis layer. The second classification only includes the occurrence of one or more signs or symptoms of infection and/or inflammation, ischaemia, osteomyelitis. The third classification includes tissue damage involving layers of the lower dermis (subcutaneous), extending to the tendon tissue but not the bone. The fourth classification describes the following tissue damage: subcutaneous/muscle/fascia/tendon with one or more sign of inflammation/infection/ischemia/osteomyelitis. The fifth classification describes damage to the entire skin tissue extending to the bone, including tissues that has experienced both localised and extensive gangrene. The sixth classification is similar to the fifth classification with the addition of any one or more of the following signs: inflammation/infection/ischemia/osteomyelitis. A diagnosis of infection

| Ulcer grading | Description |
|---|---|
| **TABLE 1. The SHID wound classification.** ||
| 1 | Epidermis and/or to dermis |
| 2 | Epidermis and/or to dermis with any one or more sign of infection/ischemic/osteomyelitis (X-ray) |
| 3 | Subcutaneous/fascia/muscle/tendon |
| 4 | Subcutaneous/fascia/muscle/tendon with any one or more sign of infection/ischemia/osteomyelitis (with X-ray) |
| 5 | Subcutaneous/fascia/muscle/tendon/joint-capsule/Bone |
| 6 | Subcutaneous/fascia/muscle/tendon/joint- capsule/Bone/with any one or more sign of infection/ischemia/osteomyelitis (with X-ray) |

was based on signs and symptoms in and around the local wound bed, the deeper structures and the surrounding skin, including soft tissue (e.g. cellulitis) [13]. The SHID also described signs of local wound infection, with erythema or redness (in the skin surrounding the wound and/or extending infection, such cellulitis), swelling, warmth, increasing malodour, oedema, purulence and new or increasing pain being classic signs of infection in any body organ [14]. Assessing wounds for clinical signs and symptoms of inflammation and infection is of particular importance in individuals with diabetes [15]. Ischaemia can manifest with accordingly, absent or weak popliteal or posterior tibial pulses; thinned or shiny skin; lack of hair on the lower leg and foot; pallor, coldness and redness of the affected area when the legs are dependent or 'dangled,' and pallor when the foot is elevated [16].

## Methods

A cross-sectional pilot study assessing the inter-rater reliability of the SHID wound classification system was conducted to assess 15 actual wounds in real-life patients visiting the Kitamura Wound Clinic in Pontianak city, Indonesia. Inter-rater reliability was also determined using 25 images of diabetic wounds obtained from medical records.

A total of 15 adult outpatients with diabetic ulcers were requested to participate in the study. The inclusion criteria were those who were physically able to participate and whose diabetic ulcers were of various sizes. A pre-study sample size calculation indicated that with 15 and 25 subjects were required for a one-tailed test, kappa ($p \geq 0.00$) and 90% power [17]. Six raters who classified diabetic ulcers based on three wound classification systems participated in this study. The raters comprised of two experienced wound care nurses (WCNs) with a minimum experience of 3 years and four expert nurses experienced in diabetic wound care. The WCNs

had previous training with wound classification systems Wagner [3] and UT [18]. The two WCNs used three wound classification systems (Wagner, UT and SHID) to assess 25 images of diabetic wounds. Two of the expert nurses assessed 15 actual wounds in real-life patients, while the two other expert nurses assessed 15 wounds images selected from 25 images used by the WCNs.

## Statistical analysis

Inter-rater reliability was expressed in terms of Cohen's kappa coefficient, which measures the agreement of scores measured by two raters. For example, a value of 0.60 denotes an acceptable agreement between assessors, whereas a value of 0.80 denotes satisfactory or good agreement. The following scale was utilised to determine agreement: score of 0.00-0.20=slight, 0.21-0.40=fair, 0.41-0.60=moderate, 0.61-0.80=substantial and 0.81-1.00=almost perfect [17]. Cohen's kappa coefficients were calculated using MedCalc® version 15.8.

## Study protocol

All 15 patients included herein were present at the outpatient clinic during the data collection period and were subcounty recruited by the WCN. Patients were informed regarding the study goals and procedures, after which they agreed to participate in the study and signed the consent forms. Thereafter, two WCNs at the Kitamura Wound Clinic assessed 25 images of diabetic wounds obtained from medical records, with two expert nurses assessing 15 of such images. The remaining two expert nurses assessed 15 actual wounds in real-life patients. The other WCNs who did not participate in this study changed the dressings. Each assessment was performed simultaneously and independently. Participants were blinded to the ratings of the other evaluators.

| TABLE 2. Inter-reliability results for each wound classification system. | | | | | |
|---|---|---|---|---|---|
| **Wound classification** | **Weighted kappa** | | | **Confidence interval (95%)** | | |
| | **RN, n (25) ^** | **EN, n (15) ^^** | **EN, n (14) ^** | **RN, n (25) ^** | **EN, n (15) ^^** | **EN, n (14) ^** |
| **Wagner** | 0.43 | 0.54 | 0.77 | 0.12-0.74 | 0.12-0.96 | 0.52-0.96 |
| **UT** | 0.34 | 0.63 | 0.5 | 0.15-0.53 | 0.30-0.97 | 0.09-0.90 |
| **SHID** | 0.81 | 1 | 0.84 | 0.65-0.97 | 1.00-1.00 | 0.61-1.00 |
| Note: ^ =photos, ^^ = actual wound | | | | | | |

## Results

This study obtained and assessed 40 images of diabetic wounds from the Kitamura Wound Clinic. However, one image could not be assessed due to poor quality. The inter-rater reliability among the WCNs who assessed wound images was moderate for Wagner (0.43; 95% CI 0.12-0.74), fair for UT (0.34; 95% CI 0.15-0.53) and almost perfect for SHID (0.81; 95% CI 0.65-0.97). Among the expert nurses who assessed wound images, the inter-rater reliability was substantial for Wagner (0.77; 95% CI 0.52-0.96), moderate for UT (0.50; 95% CI 0.09-0.90) and almost perfect for SHID (0.84; 95% CI 0.61-1.00) **(TABLE 1)**. For actual wound assessment, a total of 15 patients with a mean age of 59.30 years (standard deviation 10.99 years; range 28-82 years; 53.3% women) were enrolled herein. Accordingly, the inter-rater reliability among expert nurses who assessed actual wounds was moderate for Wagner (0.54; 95% CI 0.12-0.96), substantial for UT (0.63; 95% CI 0.30-0.97) and perfect for SHID (1.00; 95% CI 1.00-1.00) **(TABLE 2)**.

## Discussion

The current study demonstrated that the SHID wound classification system had higher inter-rater reliability compared to the Wagner and UT classification systems among three groups of raters (one group of WCNs and two groups of expert nurses). Interestingly, the overall agreement in inter-rater reliability scores between the three groups of raters using the SHID was almost perfect to perfect. The Wagner and UT classification systems showed substantial results among expert nurses who assessed actual wounds and wound images, indicating that expert nurses are better than WNs nurses in terms of wound classification using the Wagner and UT systems on wound images. The current study also showed that educational background and length of clinical experience could greatly influence the results of reliability studies. Another interesting finding is that the of the between expert nurses who

used the Wagner system to assess wound images had better reliability study results compared to those using who assessed actual wounds, with the opposite having been observed for those using the UT system. This study demonstrated that expert nurses using the Wagner system had better reliability when assessing images than when assessing real-life wounds. We surmise that using images for wound assessment can be limited due to the difficulty in accurately assessing variables, such as depth and presence of undermining or tunnelling. The current study confirmed the results of previous studies showing that the Wagner and UT systems had slight to moderate reliability [7,8,19]. However, our reliability study showed that when assessing actual wounds, UT and Wagner systems had substantial and moderate reliability, respectively. The findings presented herein still show unsatisfactory results despite expert nurses having slightly higher inter-observer reliability results with the Wagner and UT systems than those included in previous studies [7,8,19]. The current study found that the UT system had higher reliability than the Wagner system among expert nurses assessing actual wounds, suggesting that perhaps this classification may be more suitable for research purpose [20]. Overall, agreements between the nurses' inter-rater reliability ratings were fair to substantial for the UT system and moderate to substantial for the Wagner system. One previous study reported that the Wagner and UT systems are valid tools for classifying diabetic foot wounds but not when scored by single observes [8]. However, the current findings suggested that the Wagner and UT systems had insufficient inter-rater reliability among the group of WCNs, although the UT classification may still be considered more complicated than the SHID among WCNs. While the Wagner scale is widely used and easy applicable, the current study indicate that it lacked reliability. Although the UT classification system has often been used, it has been considered the most complicated system [21] that is very difficult to remember and apply in daily practice [6,8]. The SHID classification had better reliability than the Wagner and UT

systems while allowing for easier assessments given that each stage and/or layer of wounded skin exhibiting a problem is accompanied by signs and/or symptoms, such as inflammation, infection and ischaemia.

To be useful, a classification system must be easy to implement and robust enough to reliably classify all ulcers. In this study, the wound condition, clinical setting, examiner error, education background and level of clinical experience might have affected the ability of the three groups of raters to obtain different results when performing a clinical test, especially when using the Wagner and UT systems. Regarding item classification, examiner error could have occurred when wound conditions were identified and/or interpreted inconsistently. While other wound classifications systems, such as the SINBAD system, have included items on neuropathy, the SHID system does not considering that neuropathy can occur at all wound classification levels, from grades one to six. Neuropathy can also occur in patients who have no ulcers, which may complicate the identification of wounds by practitioners when included. Moreover, the SHID can evaluate each tissue type, thereby allowing for the assessment of wound progression. Prior to their application in practice, all assessment tools need to demonstrate reliability. As with other tools developed to assess wound classification, the SHID classification may be used by practitioners to prevent complications, particularly during diabetic ulcer management. Using the SHID classification for wound assessment has advantages but limitations as well. The responsiveness of the SHID classification, which indicates its ability to adequately detect changes in the appearance of diabetic ulcers over time and distinguish between healing and non-healing wounds, has yet to be evaluated.

## Conclusion

This cross-sectional pilot study showed that the SHID classification system had almost perfect inter-rater reliability for assessing diabetic ulcers. These findings provide evidence supporting the application of the SHID classification system in identifying ulcers, especially in patients with diabetes. Further studies with more participants are nevertheless required to determine the validity of our study and the utility of the SHID classification over time.

## Competing and conflicting Interest

No conflict of interest has been declared by the authors.

## References

Parisi MCR, Zantut-Wittmann DE, Pavin EJ, et al. Comparison of three systems of classification in predicting the outcome of diabetic foot ulcers in a Brazilian population. Eur J Endocrinol. 159: 417-422 (2008).

Zhang P, Lu J, Jing Y, et al. Global epidemiology of diabetic foot ulceration: a systematic review and meta-analysis. Ann Med. 49: 106-116 (2017).

Pemayun TGD, Naibaho RM. Clinical profile and outcome of diabetic foot ulcer, a view from tertiary care hospital in Semarang, Indonesia. Diabet Foot Ankle. 8: 1312974 (2017).

Monteiro-Soares M, Boyko EJ, Jeffcoate W, et al. Diabetic foot ulcer classifications: A critical review. Diabetes Metab Res Rev. 36: e3272 (2020).

Tokuda T, Hirano K, Sakamoto Y, et al. Use of the Wound, ischemia, foot Infection classification system in hemodialysis patients after endovascular treatment for critical limb ischemia. J Vasc Surg. 67: 1762-1768 (2018).

Game F. Classification of diabetic foot ulcers. Diabetes Metab Res Rev. 36: 186-194 (2016).

Bravo-Molina A, Linares-Palomino JP, Vera-Arroyo B, et al. Inter-observer agreement of the Wagner, University of Texas and PEDIS classification systems for the diabetic foot syndrome. Foot Ankle Surg. 24: 60-64 (2018).

Forsythe RO, Ozdemir BA, Chemla ES, et al. Interobserver Reliability of three validated scoring systems in the Interobserversessment of diabetic foot ulcers. Int J Low Extrem Wounds. 15: 213-219 (2016).

Camilleri A, Gatt A, Formosa C. Inter-rater reliability of four validated diabetic foot ulcer classification systems. J Tissue Viability. 29: 284-290 (2020).

Szmidt E, Kacprzyk J. The spearman and Kendall rank correlation coefficients between intuitionistic fuzzy sets. Proceedings of the 7th conference Eur soc fuzzy log technique EUSFLAT 2011 French Days Fuzzy Log Appl LFA 2011. 1: 521-528 (2011).

Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med. 18: 91-93 (2018).

Houghton P. Wound assesment tools. What does "validated" mean? Wound Care Canada. 16: 58-64 (2018).

Sibbald RG, Woo K, Ayello E. Increased bacterial burden and infection: NERDS and STONES. Adv Skin Wound Care. 3: 25-46 (2006).

Haesler E, Ousey K. Evolution of the wound infection continuum. Wounds Int 2018. 9: 6-10 (2018).

Gardner SE, Hillis SL, Frantz RA. Clinical signs of infection in diabetic foot ulcers with high microbial load. Biol Res Nurs. 12: 119-128 (2009).

Armstrong DG, Lavery LA. Diabetic foot ulcers: Prevention, diagnosis and classification. Am Fam Physician. 57: 1325-1332, 1337-1338 (1998).

Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. Phys Ther. 85: 257-268 (2005).

Rodgers A, Scott E, Welsh L. Comparing wound classification systems: Impact on diabetic heel ulceration. Wounds UK. 9: 14-19 (2013).

Santema TB, Lenselink EA, Balm R, et al. Comparing the meggitt-wagner and the university of texas wound classification systems for diabetic foot ulcers: Inter-observer analyses. Int Wound J. 13: 1137-1141 (2016).

Nather A, Jun WY, Ning T, et al. Choosing a classification system for the management of patients with diabetic foot problems. Orthop Surg Traumatol. 1: 104-110 (2017).

Smith RG. Validation of Wagner's classification: a literature review. Ostomy Wound Manag. 49: 54-62. (2003).