Assessment of wound healing: validity, reliability and sensitivity of available instruments

Pillen H, Miller M, Thomas J, Puckridge P, Sandison S & Spark JI

Abstract

Objectives: If wound assessment instruments are to be used in the periodic assessment of wound healing, they must prove to be valid, reliable and sensitive measures of wound healing. Thus, this systematic literature review aims to examine available wound healing instruments in terms of these parameters.

Method: Only instruments able to measure changes in wound healing were included in this review and not those used to predict healing, classify wounds, or measure wound characteristics per se. All wound types were suitable for inclusion.

Results: A total of 20 articles were found, evaluating the validity of 10 instruments used to monitor wound healing. No instrument satisfied all criteria required for instrument validation. Instruments used to assess pressure ulcers, notably the Pressure Ulcer Scale for Healing (PUSH) and Pressure Sore Status Tool (PSST), had been validated to the greatest extent, whilst those describing healing in leg ulcers and general or surgical wounds tended to lack comprehensive and quality evaluation.

Conclusion: This review identified substantial gaps in the literature with regard to validation of existing wound healing instruments. Future studies are needed to comprehensively validate these instruments.

Introduction

Australia is currently experiencing an increase in the ageing population. With ageing, the skin becomes thinner, more fragile and more susceptible to injury. Furthermore, pathological changes with ageing increase the risk of chronic health problems that may lead to open wounds and delayed healing. Diabetes mellitus is also a major risk factor for the development of chronic wounds and is likely to become more of an issue given its increasing prevalence in Australia and around the world. Management of chronic wounds is bound to become a significant healthcare issue into the future. Financially chronic wounds represent a major drain on resources and reportedly cost the NHS around £1 billion per year.

An important part of chronic wound management involves the periodic assessment of wound healing, which is necessary to assess the healing trajectory and recommend continued or modified treatment. Instruments to assess wound healing can help to enhance communication among clinicians by defining a common language and standardising assessment of wound characteristics. If such tools are to be used, they must prove to be valid, reliable and sensitive measures of wound healing. Thus, the aim of this review is to evaluate available wound healing instruments in terms of validity, reliability and sensitivity to change.

Only instruments able to measure changes in wound healing were included in this review and not those used to predict healing, classify wounds, or measure wound characteristics per se. All wound types were suitable for inclusion.

With respect to instrument validation, key terms used in this review are defined in Figure 1.

Evidence in the literature

The following search strategy was used to identify relevant studies.
Databases searched
1. Ovid MEDLINE In-Process and Other Non-Indexed Citations and Ovid MEDLINE (1950-present).
2. TRIP database (1997-present).
3. Cochrane library.
4. Patient-reported health instruments.
5. BMJ clinical evidence.

• **Instrument validation**: the degree to which the instrument satisfies the terms listed below.
• **Content validity**: the extent to which an empirical measurement reflects a specific domain of content.
• **Criterion validity**: the extent of the relationship between instrument measures and a more accurate measure of the phenomenon of interest.
• **Concurrent validity**: a form of criterion validity where both measures are obtained at the same time.
• **Predictive validity**: a form of criterion validity where the criterion measure is obtained in the future.
• **Intra-rater reliability**: the extent to which a single rater obtains similar ratings on subsequent testing.
• **Inter-rater reliability**: the extent to which two or more raters obtain similar ratings when measuring the same thing.
• **Sensitivity**: the ability to detect changes over time not due to measurement error.


Figure 1. Definition of terms used to describe validation of wound healing instruments.

Search strategy
1. “wound assessment tool*”
2. “wound assessment instrument*”
3. “wound healing” and (tool* or measure* or scale* or instrument*)
4. “wound assessment” and (tool* or measure* or scale* or instrument*)
5. ulcer and (tool* or measure* or scale* or instrument*)
6. burn and (tool* or measure* or scale* or instrument*)
7. amputation and (tool* or measure* or scale* or instrument*)
8. “surgical wound*” and (tool* or measure* or scale* or instrument*)
In addition, citations in all relevant publications were searched for further studies. From this, a total of 20 articles were found, evaluating the validity of 10 instruments used to monitor wound healing. These instruments only examined healing in pressure ulcers, leg ulcers and surgical wounds. The majority of articles were published in journals specialising in skin and wound care, spanning from 1986 to 2008 and all studies were conducted within the United States.

**Assessment of healing for pressure ulcers**

**Pressure Ulcer Scale for Healing (PUSH)**

The PUSH tool was designed to track pressure ulcer healing by monitoring wound parameters of length times width, exudate amount and tissue type.9

Content validity was reported to have been established from a review of the literature and from expert opinion, yet no details of methodology were provided. Hence the authors cannot be confident that bias was minimised.

Concurrent validity was evaluated by comparing total PUSH scores with total Pressure Sore Status Tool (PSST) scores and ulcer surface area over a period of 5 weeks.10 Total PUSH score was found to be strongly correlated with these concurrent measures \((p<0.001\) for all 5 weeks), with Pearson’s \(r\) increasing over the assessment period. However, the study reported exudate and tissue type items as having little effect on total PUSH score, meaning that concurrent validity may have been more sensitive to changes in ulcer size than for the other two PUSH items.

Predictive validity was evaluated by validation of the statistical model of healing. Using data from a research database of 37 pressure ulcers, Thomas et al.9 found that surface area, exudate amount and surface appearance defined the best model of healing \((p<0.01)\), with similar results found using a separate database of 10 ulcers (data not reported). Following pilot testing and revision of the PUSH, principal component analysis of 269 ulcers again found these parameters to provide the best model for healing, accounting for 39% to 57% of variation in healing.11 Therefore, evidence is consistent in suggesting predictive validity of the PUSH, strengthened by the larger sample size used in the latter study. Unfortunately, limited information was provided regarding sample characteristics, making it difficult to conclude whether claims of predictive validity are relevant to all ulcer types.

Examining sensitivity to change, Thomas et al.9 compared biweekly PUSH scores from 13 healed ulcers over 8 weeks. Significant differences \((p<.05)\) were found between scores at baseline and subsequent weeks and between week 2 and weeks 4 and 6. Differences at later intervals were non-significant, possibly the result of a decreasing number of subjects as wounds had healed (i.e. type 2 error). Similar findings were reported by Stotts et al.11 in studies involving two separate patient databases, suggesting that the PUSH tool is sensitive to changes early in the wound healing process, but not so for later stages. Given the larger sample size used in this study \((n=103\) and 269 for each database), it is less likely that sample dropout had caused these non-significant effects.

In terms of inter-rater reliability, studies have reported greater than 95%11 and 90%10 agreement between raters, yet no information was provided as to how these values were obtained. Intra-rater reliability has not been reported.
In summary, predictive validity of the PUSH tool has been demonstrated consistently in multiple studies; however, concurrent validity and intra- and inter-rater reliability still need to be investigated further. Revisions may also be necessary to improve sensitivity of the PUSH; with evidence suggesting the tool may lack sensitivity as the ulcer approaches complete healing.

**PSST**

The PSST was designed to describe wound healing in pressure ulcers, consisting of 15 scored and two non-scored items. The scored items assessed variables of wound size and depth, tissue characteristics and wound exudate, whereas the non-scored items examined wound location and shape. To date, studies have explored the tool’s content validity, concurrent validity and intra- or inter-rater reliability.

Content validity was established by a panel of nine nurses recognised as experts in the area of wound care, expressed as a content validity index value (CVI=0.91; representing the proportion of experts rating each item as quite or very relevant on a four-point scale). This suggests that the PSST is content valid, given that a CVI value ≥0.78 was required to establish content validity at the \( p=0.05 \) level.

Concurrent validity was only evaluated for the non-scored PSST item of depth, compared with classification of the ulcer’s AHCPR (Agency for Health Care Policy and Research) stage. Given that the depth item is only a small and non-scored part of the tool, this result has little relevance to the concurrent validity of the PSST. A better evaluation of concurrent validity came from Bates-Jensen, comparing PSST total score to NPUAP (National Pressure Ulcer Advisory Panel) ulcer stage, giving a correlation of \( r=0.606 \) (\( p=0.001 \)). Nonetheless, both studies are limited in the use of ulcer stage as a concurrent measure of wound healing, which may not be appropriate given the physiological inaccuracies associated with reverse-staging.

One study reported strong intra-rater reliability for total PSST score using two wound expert raters (Pearson’s \( r=0.99 \) for rater 1 and \( r=0.96 \) for rater 2, \( p<.001 \)). However, these values are likely to have been inflated due to additional training that was performed after assessing 13 of 20 ulcers. Furthermore, it is difficult to generalise results given the use...
of experienced wound nurses and the limited number of wounds (20 ulcers) assessed. In response to this, Bates-Jensen & McNees \(^{13}\) evaluated intra-rater reliability using 15 general health practitioners who did not have extraordinary training in wound care. Intra-rater reliability (using Cohen’s Kappa) for the practitioners averaged 0.89, which was considered within an acceptable range. Nonetheless, generalisability is still limited by the small number of wounds assessed (16 ulcers).

Bates-Jensen \(\text{et al.}^{14}\), reported strong inter-rater reliability for total PSST score (0.91 for time 1 and 0.92 for time 2, \(p<.001\)), but again this was limited by the use of only two raters, both of which being wound care experts. In a later study, Bates-Jensen & McNees \(^{15}\) reported acceptable inter-rater reliability for general healthcare practitioners (Kappa=0.78) and for practitioners versus an expert enterostomal nurse (Kappa=0.82). Generalisability of these results is strengthened by the use of a more diverse group of raters.

In summary, available evidence suggests that the PSST is content valid and has acceptable reliability when used by practitioners with a range of experience in wound care. However, there is uncertainty regarding concurrent validity and no attempt to establish predictive validity or sensitivity of the PSST.

**Sessing Scale (SS)**

The SS is a seven-stage scale designed to measure progress in wound healing over time, with each stage describing wound tissue attributes throughout the wound healing process \(^{16}\). To date, content, concurrent and predictive validity and intra- and inter-rater reliability have been reported in two articles \(^{16,17}\).

Content validity was evaluated by a panel of five wound care nurses. One-hundred per cent agreement was reached amongst these experts with regard to the tool’s conceptual framework, content and item hierarchy, suggesting that the SS is content valid \(^{16}\). Unfortunately, no information was provided as to how this agreement was reached.

Concurrent validity was determined by comparing scores from the SS with concurrent measures of ulcer Shea Scale score (used to classify the degree of tissue damage in pressure ulcers) and wound diameter. A strong relationship (Spearman’s \(r=0.90, p<0.0001\)) was found between changes in SS versus Shea Scale scores and a moderate to strong relationship (\(r=0.65, p<0.0001\)) was found between changes in the SS score versus wound diameter \(^{16}\). However, it is questionable whether wound diameter and Shea Scale are appropriate concurrent measures, given their limitations in monitoring wound healing \(^{16}\).

Predictive validity was evaluated by Ferrell \(\text{et al.}^{17}\) by stepwise regression analysis, finding that initial SS was more predictive of healing rate than either initial wound diameter or Shea stage (data not reported). A limitation of this method is that the SS score was identified by backwards elimination from several potential modifiers of wound healing, making it difficult to account for confounding factors.

Intra-rater reliability was evaluated on two occasions by Ferrell \(\text{et al.}^{16}\), once with 10 pressure ulcers evaluated on 2 consecutive days (pilot data: weighted kappa=0.90) and a second time with 50 ulcers (longitudinal data: weighted kappa=0.84). Both statistics demonstrated adequate intra-rater reliability, strengthened by agreement between pilot and longitudinal data. Inter-rater reliability was also evaluated in the pilot study, with two nurses independently rating 10 ulcers to give a weighted kappa of 0.80. Generalisability of this is limited by the use of only two raters and the assessment of a small number of pressure ulcers.

Although these results suggest that the SS is a valid and reliable tool for monitoring progression of wound healing, validation has only been tested using a single sample, which exclusively featured pressure ulcers of the trunk or trochanters. This limits the ability to generalise results to other ulcer types and healthcare settings. All validation work has been performed using wound care nurses, so it is uncertain how the SS would rate when used by nurses or physicians with less experience in wound care. Sensitivity of the tool is yet to be evaluated.

**Sussman Wound Healing Tool (SWHT)**

The SWHT is based on an acute model of wound healing, which describes tissue status and size throughout the wound healing process. The tool itself consists of 21 items, gathering data on wound attributes, location, healing phase and extent of tissue damage. Since being introduced in 1997 by Sussman and Swanson \(^{18}\), a scoring system has yet to be developed for the tool.

Only one study has attempted to evaluate the use of the SWHT, seeking to establish predictive validity of the tool \(^{18}\). However, only the first 10 items (tissue type and wound attributes) of the SWHT were accounted for in the analysis, likely because of the absence of a scoring system for the remaining items. While results of the study suggest that the initial SWHT score may have a role in predicting whether or not a pressure ulcer is likely to heal, this is not relevant to the monitoring of wound healing.

Until a scoring system is developed and until validity, reliability and sensitivity has been established, the SWHT...
cannot be considered an appropriate measure of wound healing.

DESIGN
DESIGN is an acronym representing six items used to describe wound-healing progression, with each item scored according to the severity or extent of that characteristic. Despite having been used in previous research, only one study has attempted to validate the DESIGN tool.

Concurrent validity was evaluated by comparing scores from the DESIGN tool and the PSST from examination of eight wound photographs. This was assessed by seven nurse raters, presenting with various levels of experience in wound care. All seven raters showed a correlation (Spearman’s $r$) $\geq 0.9$ when comparing DESIGN and PSST scores, suggestive of concurrent validity. However, this is limited by the use of photographs to assess wounds, making it difficult to accurately assess variables of exudate and undermining or tunnelling.

Inter-rater reliability was also evaluated, with seven raters used to assess eight wound photographs and six real-life wounds. Using Pearson’s correlation, the agreement rate of scores for all of the raters had a correlation of $r=0.98$ for the wound photographs and $r=0.91$ for the real-life wound assessments, both of which were high. However, results should be interpreted with caution given the small number of ulcers examined, the non-random selection of photographs or patients and the absence of details regarding ulcer characteristics.

Given the strong correlations reported for concurrent validity and inter-rater reliability, the DESIGN tool shows promise in monitoring wound healing; however, further studies with larger sample sizes are needed to confirm such findings. Further work is also required to establish intra-rater reliability and sensitivity to change.

Wound Healing Scale (WHS)
The WHS was developed as an alternative to the reverse staging of pressure ulcers, acknowledging that this method of ulcer assessment is physiologically inaccurate. The scale is based upon the existing NPUAP staging system, but with eight alphabetic modifiers that reflect the physiological changes that occur with healing. At the time of this review, no studies have attempted to validate this tool.
Assessment of healing for leg and venous ulcers (PUSH)

Ratliff & Rodeheaver \(^{21}\) evaluated the use of the PUSH tool to assess healing of venous leg ulcers, based on the opinion that parameters used in the PUSH tool may also be relevant to the assessment of leg ulcers. To date, only one study has attempted to validate the PUSH tool in this patient population \(^{21}\). Despite the author’s conclusion that the PUSH tool fulfills the “need for a simple, valid, reliable and practical tool for monitoring the process of venous ulcer healing”, there is limited evidence to support this statement. Although the suitability of PUSH items to assess healing in leg ulcers was based on a review of the literature, no attempt was made to establish content validity. The study did not evaluate criterion validity or intra-rater reliability and determined inter-rater reliability by comparing two raters, for which only five assessments were performed (giving 100% agreement). Also, the study described sensitivity without use of statistical methods, providing no indication of whether observed effects were significant. Therefore, there is no evidence to suggest that the PUSH tool is a valid measure of healing of venous leg ulcers.

Leg Ulcer Measurement Tool (LUMT)

This instrument was developed to describe changes in leg ulcer (including diabetic, venous and arterial ulcers) status over time. To date, only one study has attempted to validate the LUMT, evaluating this tool with respect to construct and concurrent validity, inter- and intra-rater reliability and sensitivity to change \(^{22}\). Construct validity was established by a panel of nine wound care specialists, reaching consensus that all suitable domains had been included and all responses were appropriate. However, no details were provided as to how consensus was reached and how bias was minimised.

Concurrent validity was evaluated by comparing total LUMT and LUMT size item scores with measures of ulcer surface area, giving Pearson’s \(r\) of 0.43 and 0.82, respectively. Since the authors required a correlation >0.75 to demonstrate concurrent validity, total LUMT score cannot be said to show concurrent validity when compared to wound area. Perhaps a more comprehensive method of wound assessment, including factors other than size, is needed to better evaluate concurrent validity.

Intra-rater reliability was reported as excellent (intraclass correlation coefficient, ICC, =0.96 for experienced and inexperienced raters), although it is possible that this result was inflated by recall bias due to repeat measures being conducted in the same half-day. Inter-rater reliability was acceptable for both experienced (ICC=0.77) and inexperienced (ICC=0.89) raters, with a higher correlation amongst inexperienced raters possibly due to additional training they received or chance variation because of the small number of raters (n=4 experienced and 2 inexperienced raters).

When wounds were followed up monthly over a period of 4 months, total LUMT score was shown to be sensitive to changes in wound status after dividing the patient group in healers, non-healers and those with no change (responsiveness coefficient= 0.84). A criticism of this method is that wounds were assessed only monthly, which may be too infrequent for use in monitoring outcomes in research or clinical settings.

In summary, the LUMT appears to show adequate intra- and inter-rater reliability and is sensitive to monthly changes in ulcer healing. However, there is some uncertainty about the tool’s concurrent validity, notably because the LUMT was not compared with a measure of wound healing able to account for non-size-related measures of ulcer healing.

Assessment of surgical and general wounds

Barber Measuring Tool (BMT)

This tool is used to track changes in wound volume over time, with wound progression presented as “percentage healed” over time \(^{23}\). At the time of this review, only one article had been published regarding use of this tool \(^{23}\). In this article, the author used the BMT to track the progress of over 400 wounds, including surgical, burn and chronic wounds (i.e. pressure ulcers). While the author reported the tool to appropriately represent healing progression, there was no data to support this claim.

Saying this, several studies support the idea that percentage wound area reduction is an important indication for differentiating between healing and non-healing wounds \(^{24}\), suggesting that the BMT may have a role in the assessment of wound healing. Still, it should be noted that these studies examined healing of ulcers, which may not be relevant to the healing of surgical or burn wounds.

ASEPSIS

This tool was originally intended for use in evaluating the efficacy of different antibiotic regimens by describing characteristics of abnormal wound healing \(^{25}\). Although the ASEPSIS tool is primarily concerned with assessment of healing with respect to postoperative wound infection, its ability to assess the clinical appearance of a wound warranted its inclusion in this review.
Hydrotul
Integrating wound care knowledge and technology for effective wound management

Hydrotul is a hydroactive, non-occlusive tulle dressing, impregnated with a non-medicated triglyceride ointment, which contains carboxymethyl cellulose (CMC) for moist wound healing.

- **Open weaved honeycomb structure** prevents the accumulation of exudate and reduces the risk of maceration to the wound and surrounding area.
- **CMC particles** have been specifically designed to attract and bind with water and/or proteins. These particles store and retain the exudate forming a gel at the wound surface to create a moist environment for accelerated wound healing.
- **Low adherent, gel forming** wound contact layer designed to prevent maceration and to reduce pain and trauma for patients during and between dressing changes.
- **Cost and time effective** Hydrotul has high absorbency levels and can stay in situ for 5-7 days to encourage undisturbed wound healing therefore, providing a saving on product costs, treatment time and physical labour.

### Applications
- Abrasions
- Skin tears where sub-cutaneous tissue is exposed
- Venous leg ulcers
- Arterial ulcers
- Decubitus ulcers

For further information or to receive a Hydrotul information kit call **1300 720 983**.

Inter-rater reliability of ASEPSIS has been evaluated by two separate studies by Wilson et al. and Byrne et al., both demonstrating the ASEPSIS tool to have high inter-rater reliability. However, both studies failed to report data on sample characteristics and used only two raters in performing the wound assessments, limiting generalisability of claims of reliability.

Although several studies have used the ASEPSIS tool in the assessment of wound infection rates, none have examined the validity, intra-rater reliability, or responsiveness of the tool.

### Granulometer

The Granulometer consists of eight wound photographs, each representing a different stage of healing, placed next to each other in order of wound healing. This is based on the idea that colour and confluence of granulation tissue are the most important indicators of open-wound healing and that changes in characteristics of wound tissue are associated with healing. Only one study has examined the use of the Granulometer in the assessment of wound healing.

In this study, a strong correlation was found between the Granulometer scale and skin graft-take grade. Since graft viability was used as a proxy for wound condition, the authors concluded that the Granulometer score was associated with wound healing progression. While this suggests that the Granulometer may be appropriate for wound assessment per se, the study did not allow for monitoring wound healing over time, notably because of the lack of longitudinal measurements using the Granulometer.

Inter-rater and intra-rater reliability were both high \( p = 0.9 \) using Friedman two-way ANOVA by ranks and \( p = 1 \) using Wilcoxon’s matched pairs signed rank test, respectively, suggesting that Granulometer measures are reproducible. However, surgeons were exclusively used for reliability testing, making it difficult to generalise results to settings in which nursing staff are responsible for wound assessments. The small number of raters (three surgeons) and number of wounds assessed (20 open wounds) further limit generalisability of these findings.

In most instances, adjacent Granulometer scores were found to have a non-significant predictive ability for graft-take, meaning that the Granulometer may not be sensitive to small changes in wound status. Again, a longitudinal assessment of wound healing using the Granulometer is needed to better assess the tool’s sensitivity to change.
Given these limitations, it seems that further work is required to evaluate the validity, reliability and sensitivity of the Granulometer in patients with open wounds, using longitudinal measures where appropriate.

**Conclusion**

From evaluation of 10 different instruments proposed to measure wound healing, none were found to satisfy all criteria required for instrument validation, defined by content and criterion validity, intra- and inter-rater reliability and sensitivity to change\(^8\). Nonetheless, studies did cover aspects of instrument validation to various extents as summarised in Table 1. From a review of the literature, it is clear that instruments used to assess pressure ulcers, notably the PUSH tool and PSST, have been validated to the greatest extent. The SS has also been well evaluated, albeit only using a single sample of patients. Instruments used to describe healing in leg ulcers and general or surgical wounds, with the exception of the LUMT, lack comprehensive and quality evaluation with respect to validity, reliability and sensitivity.

In examining studies attempting to validate these instruments, several methodological shortcomings were identified. In most cases, patient samples were small and homogenous, the sampling procedure was non-random (i.e. consecutive or convenience sampling) and validation was performed using a single sample. Also, it was common to use a small number of raters to assess reliability. All these factors are likely to limit the ability to generalise study findings to other settings and patient or rater groups, a limitation that was reported frequently throughout this review. Another frequently encountered limitation was the lack of an appropriate ‘gold standard’ measure of wound healing, especially in studies examining concurrent validity. Often a concurrent measure of wound area or diameter or ulcer classification stage was used, which may be inappropriate for monitoring progress in wound healing\(^{15}\).

From this review, there are substantial gaps in the literature with regard to validation of wound healing instruments. Thus, future studies are needed to comprehensively validate these instruments where gaps exist. Such studies should make use of larger sample sizes, greater number of raters and be conducted across multiple settings in order to improve generalisability of results.

There is insufficient evidence to suggest use of any available instrument proposed to assess wound healing. The PUSH and PSST appear to show promise in the monitoring of pressure ulcer healing based on existing evidence, warranting further research to establish their validity and reliability. The LUMT also shows promise for measuring healing of leg ulcers.
although validity of the tool still needs to be established. Considerable work is still needed before instruments assessing other types of wounds can be used in practice.

Conflicts of interest

None identified.

References