ORIGINAL ARTICLE

Quality of measurements of acute surgical and traumatic wounds using a digital wound-analysing tool

Dymmie LC Landa[1], Anne-Margreet van Dishoeck[1,2], Ewout W Steyerberg[2] & Steven ER Hovius[1]

1 Department of Plastic, Reconstructive and Hand surgery, Erasmus University Medical Center, Rotterdam, the Netherlands
2 Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands

Landa DLC, van Dishoeck AM, Steyerberg EW, Hovius SER. Quality of measurements of acute surgical and traumatic wounds using a digital wound-analysing tool. Int Wound J 2016; 13:619–624

Abstract

The aim of this study was to assess the reliability of measurements using a wound-analysing tool and their interpretability. Wound surface areas and tissue types, such as granulation, slough and necrosis, in twenty digital photographs were measured using a specific software program. The ratio of these tissue types in a wound was calculated using a wound profile. We calculated the intraclass coefficient or κ for reliability, standard error of measurement (SEM) and smallest detectable change (SDC). The inter-rater reliability intraclass correlation coefficient (ICC) was 0·99 for surface area, 0·76 for granulation, 0·67 for slough and 0·22 for necrosis. The profiles gave an overall κ of 0·16. For test–retest reliability, the ICC was 0·99 for surface area, 0·81 for granulation, 0·80 for slough and 0·97 for necrosis. The agreement of the applied profiles in the test–retest was 66% (40–100). SEM and SDC for surface area were 0·10/0·27; for granulation, 6·88/19·08; for slough, 7·17/19·87; and for necrosis, 0·35/0·98, respectively. Measuring wound surface area and tissue types by means of digital photo analysis is a reliable and applicable method for monitoring wound healing in acute wounds in daily practice as well as in research.

Introduction

Wound assessment is a dynamic process providing vital information to ensure patients receive the appropriate interventions at the right time. Wound measurement is an important part of this assessment. Impaired wound healing can be recognised if measurements of wound characteristics, at least for wound surface area and tissue types, are performed (1). Repeated measurement of wound surface area can help in following wound healing over time. Tissue types provide a baseline against which wound treatment can be assessed (2,3). There are different methods for measuring wound characteristics, such as invasive or non-invasive, and objective or subjective (4–6). The measurement process in daily practice, however, has some factors that need to be given attention (7). The description of wound characteristics is usually subjective and not described in unambiguous terminology (8). Both the methods of wound measurement and the choice of descriptions of the wound characteristics are dependent on the knowledge and skills of the health care professional (9). Therefore, wound healing is difficult to monitor because the interpretations by different observers can be quite divergent (10). This hinders proper treatment at the right time and entails the consequences.

A digital wound analysis tool provides a standardised and objective method for wound measurement. This may be the answer to the issues mentioned here (11). Surface area and tissue types are calculated from digital photographs by means of a specific software program. Measurement properties such as

Key Messages

- a digital wound analysis tool provides a standardised and objective method for wound measurement
- measurement properties such as reliability and interpretability, the ability to detect clinically true changes beyond change, are important for treatment planning and for monitoring progress
- measuring wound surface area and tissue types by means of digital photography analysis is reliable and applicable in monitoring wound healing in surgical and traumatic wounds

reliability and interpretability, the ability to detect clinically true changes beyond change, are important for treatment planning, monitoring progress, and evaluating treatment response. To our knowledge, no inter-rater reliability test was available for the digital wound analysis tool.

The aims of this study were to evaluate the inter-rater reliability and test–retest reliability of this instrument as well as the accuracy of the instrument in measuring true changes beyond change. We addressed three major questions:

1 Do two observers agree when applying the same measurement on the same wound?
2 Does the same observer assess the wounds consistently in the test–retest with a 2-month interval?
3 How accurate are the changes of surface area and tissue types measured with the wound analysis tool?

Methods and materials

Study design

The observational reliability study was performed in a university hospital, the Erasmus Medical Centre in Rotterdam, The Netherlands. Reliability was tested by means of inter-rater and test–retest reliability. Accuracy was calculated using standard error of measurement (SEM) and the smallest detectable change (SDC).

Wound-analysing tool

Digital imaging and computerised planimetry have been found to be accurate in measuring wound size (12,13). We incorporated computerised planimetry in our electronic patient records. The Wound Healing Analyzing Tool (W.H.A.T.®; BAP Medical BV, Apeldoorn, the Netherlands) was developed by Wild *et al.* (11) at the Medical University of Vienna, Austria. This software calculates the wound surface area and the percentage of the tissue type's granulation, slough and necrosis in a wound. The software recognises shades of red, yellow and black of every pixel of the photo. The validity of the analyses was tested using histological samples of chronic wounds (11). After manual tracing of the wound boundary with a computer on a digital photograph (Figure 1A), a calibration is performed using a predefined square on the ruler next to the wound on the photograph. The program recognises the printed square mark placed next to the wound to evaluate the size of the wound (11). Subsequently, the observer applies one of the profiles for analysis. These profiles are based on tissue types (granulation, slough and necrosis) or wound type (abdominal, split-thickness skin and superficial skin) and guide the software's computation. As a next step, the analysis is performed, resulting in a wound image with the calculated measures of surface area (cm^2) and the percentage of tissue types (Figure 1B). Validation of size measurements was performed by comparing the measurements with those obtained using other methods of wound sizing (11).

Population

The wound photos were selected from a research database where we studied wound characteristics in acute open surgical
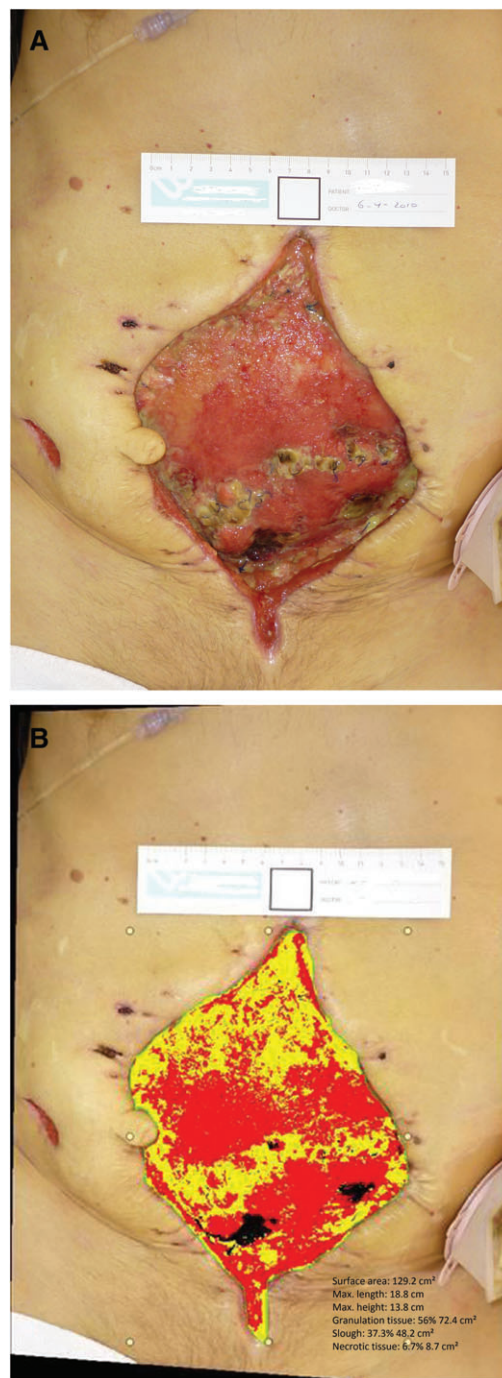

Figure 1 A digital photograph showing calibration that is performed using a predefined square on the ruler next to the wound on the photograph (A). Wound image with the calculated measures of surface area (cm^2) and the percentage of tissue types (B).

and traumatic wounds. This study was approved by the Erasmus Medical Centre Ethics Committee (MEC-2009-346). The database consists of 457 digital wound photographs, taken in 130 patients from the second half of 2009 until the end of 2010 using a standardised method. Out of these, the researcher qualitatively selected 20 photographs, which provided a large range of wounds regarding surface area and tissue type. The inclusion

Table 1 Descriptive

Patients	*N = 20*
Gender, male *n* (%)	13 (65)
Age*	49 (43–57)
Wound location *n* (%)	
Limbs	12 (60)
Thorax-abdomen	7 (35)
Buttocks-genital	1 (5)
Head and neck	0
Photographs	*N = 20*
Surface area*	22·1 (13·5–58·3)
Granulation*	56·6 (22·0–68·3)
Slough*	39·1 (27·8–76·2)
Necrosis*	2·1 (0·5–6·3)
Observers	*N = 9*
Gender, male *n* (%)	3 (33·3)
Age*	51 (39·5–58·0)
Profession *n* (%)	
Doctor	3 (33·3)
Nurse	6 (66·6)
Years of working experience*	20 (6·5–29·0)

*Values are median (interquartile range).

criteria for the pictures were differences in terms of wound type and wound location and visual differences in size and tissue types – granulation, slough and necrosis (Table 1).

Observers

A team of nine observers, including three specialised wound nurses, three nurses with experience in wound care and three plastic surgeons, were asked to analyse the 20 wound photographs individually. The observers' experience in wound care varied from 3 years up to 36 years, with a median of 20 years (Table 1). The nurses' experience in using the analysing tool varied from daily use to once a week. The plastic surgeons had no former experience in using this tool. All observers received instructions before they started analysing the wound photographs.

Measurements

The measurement outcomes of the surface area, tissue types and profiles were recorded on a data collection form. The surface area was presented in cm^2. Tissue types were presented as a percentage of the surface area. Both cm^2 and percentage were regarded as continuous variables. We considered the profiles as a nominal variable.

Inter-rater reliability is the degree to which two or more individuals agree on what they observe. The inter-rater reliability was calculated using the combined data from nine observers and 20 wound photographs. Test–retest reliability is a test of an instrument's stability, assessed by repeated measurements over time (14). The test–retest reliability was calculated using the additional data from five observers (all nurses), who measured the wounds twice, with an interval of 2 months.

Statistical analysis

Reliability was measured using the intraclass correlation coefficient (ICC) with a two-way random effects model of absolute agreement, in order to quantify the degree of agreement between the measured values on a continuous scale (surface area and tissue types) (15). To measure the agreement between the profiles, Cohen's kappa (κ) statistics was used as a parameter of reliability for nominal scales (16,17). In addition, we calculated the percentages of agreement. We used the scale of agreement defined by Landis and Koch for interpretation of the values of the ICC en κ (18). We valued κ or ICC above 0·80 as 'very good', between 0·80 and 0·60 as 'good', between 0·60 and 0·40 as 'moderate' and below 0·40 as 'poor'.

Reliable changes can be defined as 'a noticeable, appreciable difference that is of value to the health professional, and that exceeds variation attributable to chance' (19–21). To assess the measurements of this instrument for clinical practice, we used the outcomes of the ICC two-way random effects consistency and the $SD_{difference}$ of the test–retest to calculate *SEM*. The *SEM* approximates how repeated measurements of a person on the same instrument tend to be distributed around the 'true' score. The SEM is the SD around a single measurement (22). We calculated the group estimate as follows:

$$SEM = SD_{difference} \sqrt{\left(1 - ICC_{consistency}\right)}.$$

The SEM also allows the calculation of the SDC. The SDC is an estimate of the smallest change in the score that can be detected objectively for a wound, that is, the amount by which a wound's score needs to change to make sure the change is greater than a measurement error. The SDC is formulated as follows:

$$SDC = 1.96 \sqrt{2 \times SEM}.$$

All analyses were performed using SPSS PASW version 20.0. Cohen's κ was performed using SAS 9.2.

Results

The median of the surface area of the wounds used in this study was $22 \cdot 1\ cm^2$. The median for granulation was 56·6%, for slough 39·1% and for necrosis 2·1% (Table 1). The observers chose mainly the profile for partial granulation (56·7%).

The ICC for inter-rater reliability was very good (0·99). The statistics for granulation and slough showed good agreement (ICC 0·76 and 0·67) while that for necrosis was poor (ICC 0·22) (Table 2). For the profiles, we found a poor agreement (overall κ 0·16, Table 3A)

We observed differences between the groups of observers in the inter-rater assessment, particularly on tissue types. The agreement between the three nurses with experience in wound care and the plastic surgeons was the same for surface area (ICC 0·99). Granulation and necrosis, ranging from good for granulation (ICC 0·75) to poor for necrosis (ICC 0·06). The agreement for the measurement of slough was good in the assessment by the surgeons (ICC 0·62) and moderate (ICC 0·56) in the assessment by the nurses with experience. Among the three specialist wound nurses, there was little variation in the assessment of surface area and tissue types (Table 2). The ICC for surface area among these nurses was very good (ICC 0·99) and that for all tissue types was good (ICC 0·78–0·80).

Table 2 Inter-rater reliability

Raters	Surface area	Granulation	Slough	Necrosis
Specialist nurses $n = 3$ ICC (CI)	0·99 (0·99–1·0)	0·80 (0·64–0·91)	0·79 (0·61–0·90)	0·78 (0·63–0·91)
Experienced nurses $n = 3$ ICC (CI)	0·99 (0·98–1·0)	0·72 (0·52–0·87)	0·56 (0·31–0·77)	0·11 (−0·13–0·42)
Plastic surgeons $n = 3$ ICC (CI)	0·99 (0·99–1·0)	0·75 (0·54–0·89)	0·62 (0·38–0·81)	0·06 (−0·17–0·37)
All raters, $N = 9$ ICC (CI)	0·99 (0·98–1·0)	0·76 (0·62–0·88)	0·67 (0·52–0·82)	0·22 (0·09–0·42)

CI, 95% confidence interval; ICC, intraclass correlation coefficient.

Table 3 (A) Inter-rater reliability profiles. (B) Test–retest reliability profiles

(A)*

Observer	1	2	3	4	5	6	7	8	9
1									
2	0·33								
3	0·07	**0·54**							
4	0·12	0·12	0·26						
5	0·09	0·18	0·18	0·39					
6	0·04	0·06	−0·06	0·04	0·14				
7	0·14	0·26	0·37	0·08	0·00	−0·04			
8	0·14	0·32	0·13	**0·45**	0·33	0·19	0·14		
9	−0·01	−0·08	0·05	0·36	**0·46**	0·07	−0·06	0·09	

(B)

Photograph	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5	Agreement (%)
1	Yes	Yes	Yes	Yes	No	80
2	Yes	Yes	Yes	Yes	No	80
3	Yes	No	Yes	Yes	Yes	80
4	Yes	Yes	No	No	No	40
5	Yes	Yes	No	Yes	No	60
6	No	Yes	No	No	Yes	40
7	No	Yes	No	No	Yes	60
8	No	Yes	Yes	Yes	Yes	80
9	No	No	No	Yes	yes	40
10	Yes	Yes	No	No	Yes	60
11	No	Yes	No	No	Yes	40
12	Yes	Yes	Yes	Yes	Yes	100
13	Yes	Yes	No	Yes	No	60
14	Yes	Yes	Yes	Yes	Yes	100
15	No	Yes	Yes	Yes	Yes	80
16	No	Yes	Yes	Yes	No	60
17	Yes	Yes	Yes	Yes	No	80
18	Yes	Yes	Yes	Yes	Yes	100
19	Yes	No	No	Yes	Yes	40
20	No	Yes	No	Yes	No	40
Agreement (%)	60	85	50	75	60	

*Kappa statistics, bold values indicate moderate agreement. κ value above 0·80, 'very good'; between 0·80 and 0·60, 'good'; between 0·60 and 0.40, 'moderate'; below 0·40, 'poor'.

The ICC results for the test–retest reliability among five observers were very good for surface area (0·99), granulation (0·81), and necrosis (0·97). Test–retest reliability for slough was good (0·80). The same profile as in the test–retest was applied in 66% of the photographs (40%–100%) (Table 3B).

The smallest SEM score was for surface area (0·10 cm^2) and the largest SEM score was for slough (7·17%) (Table 4). The SDC for surface area was 0·27, meaning that a change of 0·27 cm^2 was detectable beyond common cause variation. For granulation, this was 19·08%, for slough 19·87% and for necrosis 0·98%.

Discussion

In this study, we evaluated the reliability and interpretability of an instrument in the objective analysis of surface area and tissue types in surgical and traumatic wounds. We demonstrated that this wound-analysing tool provides a reliable measurement of the wound surface area, measured by inter-rater agreement (ICC 0·99), as well as by intra-rater agreement (ICC 0·99). We use this tool in daily practice. Processing a photo takes several minutes but provides an objective measurement that has advantages in clinical practice as well as in research. In clinical practice,

Table 4 Standard error of measurement and smallest detectable change

Variable	Test–retest ICC$_{consistency}$	SD$_{difference}$ (test–retest)	SEM	SDC
Surface area	0·99	2·18	0·10	0·27
Granulation	0·81	15·79	6·88	19·08
Slough	0·80	15·95	7·17	19·87
Necrosis	0·97	2·07	0·35	0·98

ICC, intraclass correlation coefficient; SDC, smallest detectable change; SEM, standard error of measurement.

it enhances communication and in research, it provides an objective and accurate measure of the surface area. It offers a precise calculation of the surface area (SEM $0.10\,cm^2$), and a change in surface area of $0.27\,cm^2$ is detectable beyond random variation. Compared with other techniques of surface area measurement, such as ruler-based techniques and tracing the wound edges onto a transparent gridded film, the wound analysis tool provides a more precise measurement. Ruler-based techniques are inconsistent and not reliable in the case of irregular or large wounds, for they overestimate the wound area by up to 70%. (23) Tracing the wound edges on a transparent paper and using a metric grid to count the number of square centimetres for the estimation of the wound surface area has shown inaccuracy as a consequence of estimation of partial squares (24).

In this study, the inter-rater reliability measures vary for the different tissue types. Considering the SEM, the outcomes of the percentage granulation and slough calculated by the wound analysis tool will be higher or lower in a wound. While using this analysis tool, the professional must consider this measurement error. The outcomes of the analysis will vary by 7% for granulation, 7% for slough, and 1% for necrosis. Limited studies are available on the reliability of visually estimating a percentage range for tissue types (25). Compared with the results of research by Mekkes and Westerhof, the differences vary from 15% to 36% in visual estimation of the percentage of wound tissue types between two observers (25). The measurement error for tissue types calculated by the wound analysis tool is lower than the variation in visual estimation in this study, thus making it more reliable.

In our study, the following percentage of changes (SDC) could be detected reliably: 19% for granulation, 20% for slough and 1% for necrosis. Only a few studies examined the decrease and the increase in granulation, slough, and necrotic tissue over time (25–27). In a small study, the percentage of slough decreased 12% on day 4 and 56% on day 7 (25). In a case study, a 50% increase in granulation was seen in 7–12 days and a 10% decrease in necrotic tissue in 3 weeks (26). A 31% increase in necrotic tissue in diabetic wounds in 2 weeks' time has been described earlier (27). From these studies, it can be concluded that measuring a wound once a week is adequate, since the instrument provides reliable measurements of these changes.

The inter-rater reliability of the measurement of necrosis in our study is less reliable (ICC 0·22). This may be the result of the low percentage of necrosis in the wound photos we used in this study. We selected the wounds from our database and the median necrosis in the wounds of these 130 patients was 2·1. It could also have resulted from the experience of the

observers. We observed differences in the inter-rater reliability when necrosis was scored. Nurses with experience and surgeons scored much lower (ICC 0·06) than the wound specialists (ICC 0·78) who used the tool on a daily basis. This might indicate that experience in recognising necrosis from wound photos and assessing wound photos using this tool is essential.

We observed a poor agreement in the profiles. This could have been the result of observer preference as described in a study of Maylor (9). In this correlational study investigating the relationship between word preferences and assessor's personality type, subjective types of approach to assessment could be identified. In wounds, healing and deterioration are assessed relative to different signs or symptoms. As the observers have different medical backgrounds, the results of the wound analysis are approached from different perspectives. From the treatment perspective, the observer will focus on the worst type of tissue in the wound, and the amount of necrosis. For wound healing, observers will be focused on the healthiest type of tissue, namely granulation. In both cases, the observer will choose the outcome of the analysis that suits his / her point of view. A solution to reduce the user preference and improve the poor agreement using the profiles can be to reduce the number of profiles for these acute wounds.

The methodological limitation of this study was the use of a small sample size of only 20 selected wound photographs. More wound photographs would have meant an extra burden of time for the observers. An option could have been to enlarge the group of observers. Another limitation was the small percentage of necrosis in the wounds; this mainly had an impact on the reliability measures of necrosis. Future multicentre studies with a larger group of observers and a larger percentage of necrosis in the wounds could provide a more precise outcome of the reliability of these measurements. We tested the measurement properties on the surface area, but not on the perimeter. Not assessing the reliability of the perimeter measurement is a limitation of our research. Since both measurements are strongly related as they use the same manual tracing, we assume that the measurement properties will be comparable, but this needs further research. We did not use the tool on curved surfaces in this study. This is a limitation of digital imaging planimetry (28).

We conclude that measuring wound surface area and tissue types by means of digital photo analysis is a reliable and applicable method for monitoring wound healing in acute wounds in daily practice as well as in research.

Acknowledgements

References

1. Flanagan M. Wound measurement: can it help us to monitor progression to healing? *J Wound Care* 2003;**12**:189–94.
2. Sheehan P, Jones P, Giurini JM, Caselli A, Veves A. Percent change in wound area of diabetic foot ulcers over a 4-week period is a robust predictor of complete healing in a 12-week prospective trial. *Diabetes Care* 2003;**26**:1879–82.

3. Krasner D. Wound care: how to use the red-yellow-black system. *Am J Nurs* 1995;**95**:44–7.

4. Lait M, Smith LN. Wound management: a literature review. *J Clin Nurs* 1998;**7**:11–7.

5. Shaw J, Hughes CM, Lagan KM, Bell PM, Stevenson MR. An evaluation of three wound measurement techniques in diabetic foot wounds. *Diabetes Care* 2007;**30**:2641–2.

6. Laplaud A, Blaizot X, Gaillard C, Morice A, Lebreuilly I, Clement C, Parienti JJ, Dompmartin A. Wound debridement: comparative reliability of three methods for measuring fibrin percentage in chronic wounds. *Wound Repair Regen* 2010;**18**:13–20.

7. Mayrovitz H, Smith J, Ingram C. Comparisons of venous and diabetic plantar ulcer shape and area. *Adv Wound Care* 1998;**11**:176–83.

8. Keast DH, Bowering CK, Evans AW, Mackean GL, Burrows C, D'Souza L. MEASURE: a proposed assessment framework for developing best practice recommendations for wound assessment. *Wound Repair Regen* 2004;**12**(3 Suppl):S1–17.

9. Maylor ME. Establishing nurses' preferences in wound assessment: a concept evaluation. *J Clin Nurs* 2006;**15**:444–50.

10. Romanelli M, Miteva M, Romanelli P, Barbanera S, Dini V. Use of diagnostics in wound management. *Curr Opin Support Palliat Care* 2013;**7**:106–10.

11. Wild T, Prinz M, Fortner N, Krois W, Sahora K, Stremitzer S, Hoelzenbein T. Digital measurement and analysis of wounds based on colour segmentation. *Eur Surg* 2008;**40**:5–10.

12. Bilgin M, Gunes UY. A comparison of 3 wound measurement techniques: effects of pressure ulcer size and shape. *J Wound Ostomy Continence Nurs* 2013;**40**:590–3.

13. Kantor J, Margolis DJ. Efficacy and prognostic value of simple wound measurements. *Arch Dermatol* 1998;**134**:1571–4.

14. Melnyk BM, Fine-out-Overholt E. *Evidence-based practice in nursing & healthcare: a guide to best practice*. Philadelphia: Lippincott Williams & Wilkins, 2005.

15. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420–8.

16. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *J Am Phys Ther Assoc* 2005;**85**:257–68.

17. Abbas S, Tavakoli M, Walker R. Using macro to simplify to calculate multi-rater observation agreement. Paper PO-05, SESUG 2012:1–11. [WWW document]. URL http://analytics.ncsu.edu/sesug/2012/PO-05.pdf [accessed on 20 September 2012]

18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.

19. Eisen SV, Ranganathan G, Seal P, Spiro A 3rd.. Measuring clinically meaningful change following mental health treatment. *J Behav Health Serv Res* 2007;**34**:272–89.

20. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Med Care* 2000;**38**(9 Suppl):II84–II90.

21. Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T, Clinical Significance Consensus Meeting Group. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res* 2005;**14**:285–95.

22. Vet de HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine*. Cambridge: Cambrigde University Press, 2011.

23. 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**:46–53.

24. Sprigle S, Nemeth M, Gajjala A. Iterative design and testing of a hand-held, non-contact wound measurement device. *J Tissue Viability* 2012;**21**:17–26.

25. Mekkes J, Westerhof W. Image processing in the study of wound healing. *Clin Dermatol* 1995;**13**:401–7.

26. Schmuckler J. Acoustic pressure wound therapy to facilitate granulation tissue in sacral pressure ulcers in patients with compromised mobility: a case series. *Ostomy Wound Manage* 2008;**54**:50–3.

27. Sherman RA. Maggot therapy for treating diabetic foot ulcers unresponsive to conventional therapy. *Diabetes Care* 2003;**26**:446–51.

28. Harding KG. Methods for assessing change in ulcer status. *Adv Wound Care* 1995;**8**:37–42.