burns 45 (2019) 261–281

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/burns

Review

A systematic review on the quality of measurement techniques for the assessment of burn wound depth or healing potential

Mariëlle E.H. Jaspers [a,b,c,d,*], Ludo van Haasterecht [a], Paul P.M. van Zuijlen [a,b,c,d], Lidwine B. Mokkink [e]

[a] Burn Center, Department of Surgery, Red Cross Hospital, Beverwijk, The Netherlands
[b] Association of Dutch Burn Centers, Beverwijk, The Netherlands
[c] Department of Plastic, Reconstructive and Hand Surgery, Red Cross Hospital, Beverwijk, The Netherlands
[d] Department of Plastic, Reconstructive and Hand Surgery, VU University Medical Center, Amsterdam Movement Sciences, Amsterdam, The Netherlands
[e] Department of Epidemiology and Biostatistics and Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Accepted 17 May 2018

Keywords:
Burn wound
Depth
Healing potential
Measurement properties
Assessment
Review [publication type]

ABSTRACT

Purpose: Reliable and valid assessment of burn wound depth or healing potential is essential to treatment decision-making, to provide a prognosis, and to compare studies evaluating different treatment modalities. The aim of this review was to critically appraise, compare and summarize the quality of relevant measurement properties of techniques that aim to assess burn wound depth or healing potential.

Methods: A systematic literature search was performed using PubMed, EMBASE and Cochrane Library. Two reviewers independently evaluated the methodological quality of included articles using an adapted version of the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. A synthesis of evidence was performed to rate the measurement properties for each technique and to draw an overall conclusion on quality of the techniques.

Results: Thirty-six articles were included, evaluating various techniques, classified as (1) laser Doppler techniques; (2) thermography or thermal imaging; (3) other measurement techniques. Strong evidence was found for adequate construct validity of laser Doppler imaging (LDI). Moderate evidence was found for adequate construct validity of thermography, videomicroscopy, and spatial frequency domain imaging (SFDI). Only two studies

Abbreviations: AUC, area under the curve; COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; FOCI, fiber-optic confocal imaging; ICC, intraclass correlation coefficient; LDI, laser Doppler imaging; LDLS, laser Doppler line scanner; LoA, limits of agreement; MIC, minimally important change; NIRS, near-infrared spectroscopy; NPV, negative predictive value; OCT, optical coherence tomography; PAI, photoacoustic imaging; PPV, positive predictive value; PROM, patient reported outcome measure; PSI, pulse speckle imaging; SEM, standard error of measurement; SFDI, spatial frequency domain imaging; SIA, spectrophotometric intracutaneous analysis.

* *Corresponding author at*: Burn Center, Department of Surgery, Red Cross Hospital, Beverwijk, The Netherlands.
E-mail address: mjaspers@rkz.nl (M.E.H. Jaspers).

https://doi.org/10.1016/j.burns.2018.05.015

reported on the measurement property reliability. Furthermore, considerable variation was observed among comparator instruments.

Conclusions: Considering the evidence available, it appears that LDI is currently the most favorable technique; thereby assessing burn wound healing potential. Additional research is needed into thermography, videomicroscopy, and SFDI to evaluate their full potential. Future studies should focus on reliability and measurement error, and provide a precise description of which construct is aimed to measure.

Contents

1. Introduction . . . 262
2. Material and methods . . . 263
   2.1. Search strategy . . . 263
   2.2. Selection of studies . . . 263
   2.3. Assessment of methodological quality of studies . . . 263
   2.4. Data extraction . . . 263
   2.5. Assessment of the quality of measurement properties . . . 264
   2.6. Synthesis of evidence . . . 264
3. Results . . . 264
   3.1. Search . . . 264
   3.2. Quality of studies and measurement properties . . . 265
   3.3. Synthesis of evidence . . . 265
      3.3.1. Reliability . . . 265
      3.3.2. Hypotheses testing for construct validity . . . 265
4. Discussion . . . 269
5. Conclusions . . . 279
   Conflict of interest . . . 279
   Acknowledgements . . . 279
   References . . . 279

1. Introduction

Burn wounds are heterogeneous and continuously subject to change, which makes it difficult to determine the severity [1–3]. Nevertheless, reliable and valid assessment of burn wound severity is fundamental to clinical decision-making. For example considering timing of surgery and choosing the right treatment protocol. In addition, adequate diagnosis may shorten the time to appropriate treatment, which is obviously beneficial for patients. A superficial burn wound will heal spontaneously whereas severe or deep wounds require surgery in the form of skin grafting. Most challenging are burn wounds where the diagnosis is indeterminate from a clinical perspective. Furthermore, in research, it is essential to be able to compare studies evaluating different treatment modalities in relation to the initial severity of a burn wound.

Worldwide, clinical evaluation of burn wound severity is the most frequently used method as it is readily available [4,5]. This method is based on visual and tactile inspection of wound characteristics such as appearance, capillary refill and sensibility, and usually expressed as burn wound *depth* [6–8]. However, the validity of clinical evaluation is limited [4,5]. This is likely due to the fact that it is extremely difficult to assess the extent of tissue damage by only inspection of the surface, especially during the first days post-burn. Moreover, the reliability of this method is moderate due to the inconsistency in ratings of different clinicians [1,4,7,9]. Another method that can be used to assess burn wound depth is histological analysis of punch biopsies obtained from the burn wound. However, this method comprises disadvantages such as a lack of standardized histological interpretation [10,11], high rates of sampling error due to heterogeneity within the burn wound, biopsy site morbidity, and inter-observer variation between pathologists [12–14].

In addition to burn wound depth, other constructs are assessed to inform optimal clinical decision-making, such as blood flow, temperature, oxygen saturation, or melanin content [15–17]. The remaining blood flow after burn injury, as measured by laser Doppler imaging (LDI) [18,19], is also expressed as 'healing potential', which is the expected time to heal of the burn wound in days [5]. Although blood flow within a burn wound seems related to the depth of the burn wound [2,20], the precise relation remains unclear. Therefore, healing potential is not a direct measure of burn wound depth by itself [21]. It rather represents a predictive variable, which can potentially help clinical decision-making. Accordingly, different burn wound constructs can be acknowledged, among which burn wound depth and burn wound healing potential are the most frequently used.

To enable the use of adequate outcome measurement techniques in the future, it is important to perform research in which the measurement properties (i.e. quality) of techniques are evaluated, as it provides information on whether the measurement technique, and thus the scores produced by the technique, can be trusted. Although other systematic reviews were performed previously, these reviews did not evaluate the methodological quality of included studies and/or did not concentrate on the technique's measurement properties [6,22–24]. Therefore, the objective of this review was to critically appraise the relevant measurement properties of techniques that aim to assess burn wound depth or healing potential, and to provide a summary and best evidence synthesis on the measurement properties of each technique. Ultimately, we aimed to provide a recommendation on the most suitable technique for burn wound depth or healing potential.

2. Material and methods

2.1. Search strategy

A systematic literature search was performed using PubMed, EMBASE and Cochrane Library, supervised by a librarian of the VU University Medical Center. The search was reported according to the PRISMA guidelines [25]. Search terms on burn wounds were combined with assessment-related search terms (i.e. imaging, measuring, and monitoring), and with different aspects to be measured (i.e. depth, healing potential, blood flow, and perfusion). The full search strategy per database can be found in Supplementary Appendix A.

2.2. Selection of studies

Inclusion and exclusion criteria are described in Table 1. MJ and LvH independently performed title, abstract, and full text screening to select eligible articles. Reviewer disagreements were identified using the web-based software platform Covidence (www.covidence.org). Disagreements were resolved through discussion, and if necessary, a third reviewer (LM) was involved to reach consensus. Subsequently, reference lists of the selected articles were searched for additional studies.

2.3. Assessment of methodological quality of studies

To assess whether results obtained from the included studies can be trusted, the methodological quality of the studies was assessed using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist.[26,27] However, adaptations were needed because the COSMIN checklist was originally developed for use in studies on the quality of patient-reported outcome measures (PROMS) (Supplementary Appendix B). Adaptations comprised, for example, adding standards about whether the environment was stable when evaluating the reliability in an inter-observer reliability study, or removing items that were not relevant or applicable, such as 'were the administrations independent?' and 'was the time interval appropriate?' as static images were often evaluated. Items on sample size requirement were assessed to provide general information, but not taken into account when drawing conclusions.

For each study on a measurement property, a set of standards about design requirements and preferred statistical methods was scored on a 4-point scoring scale, using the 'worst score counts' principle to determine the methodological quality [28].

In this review, we were particularly interested in the reliability, measurement error, and construct validity of the techniques. As we did not expect to find any studies on the measurement properties structural validity, content validity and internal consistency, these measurement properties were not assessed in this review, but studies will be described when found. Moreover, there are no guidelines available on how to evaluate content validity for clinician-reported outcome measures. In addition, since no criterion (i.e. gold standard) exists in burn wound assessment, criterion validity is not relevant. Indeed, clinical evaluation and histology are often assigned as criterion; however, these methods are associated with moderate validity and therefore not considered as an appropriate gold standard. Cross-cultural validity was not applicable as it concerns translation of the items of a PROM. Responsiveness was omitted as it detects change over time in the construct to be measured, whereas assessment of burn wound depth or healing potential is regularly performed for diagnostic purposes at one time point.

MJ and LvH independently scored the methodological quality of the included studies. Disagreement was resolved through discussion or with involvement of a third reviewer (LM). We used the COSMIN taxonomy to decide which measurement property was evaluated in a study [29]. When a single article comprised multiple studies, each study was assessed and rated separately.

2.4. Data extraction

From all included articles, data were directly extracted into Excel tables concerning: (1) general characteristics of the study and its population (i.e. animal/human study, number of burns, year, and country); (2) general characteristics of the

Table 1 – Inclusion and exclusion criteria.

Inclusion criteria	1. Full text articles published before July 2016 in English or Dutch; 2. The aim of the study was to evaluate one or more measurement properties of a measurement technique (a 'tool') for the assessment of burn wound depth or healing potential; 3. The study population consisted of humans (of all ages) or animals
Exclusion criteria	1. The aim of the study was to evaluate clinical evaluation and/or histology as measurement technique (no 'tool' is used), and/or photographic assessment (concerning a derivative of clinical evaluation); 2. Studies that focused on burn *scars*

measurement techniques; and (3) results of the evaluated measurement properties.

2.5. Assessment of the quality of measurement properties

For each study, MJ evaluated the quality of the measurement properties using predefined criteria as presented in Table 2. These criteria were adapted from the proposed criteria for good measurement properties by Prinsen et al. [30]. The possible ratings for a measurement property were adequate (+), indeterminate (?), or inadequate (−).

To evaluate hypotheses testing for construct validity, hypotheses need to be formulated a priori. Since many authors failed to formulate hypotheses a priori, we defined hypotheses about expected results between the measurement tool under study and comparator instrument that are frequently used in burn wound assessment (i.e. clinical evaluation, histological analysis, observed healing, and healing potential as measured by LDI):

- The correlation between the measurement tool and the comparator instrument is >0.7
- The specificity of the measurement tool for distinguishing a deep burn wound (no healing) from a superficial burn wound is 70%
- The positive predictive value (PPV) of the measurement tool is 70%
- The area under the curve (AUC) for distinguishing deep burn wounds (no healing) from superficial burn wounds (spontaneous healing) is >0.7

The more hypotheses are being tested on a specific measurement technique, the more evidence is gathered. It is considered here that high specificity and PPV are deemed more important than sensitivity and negative predictive value (NPV), as it is essential to avoid burn wounds that are falsely classified as needing surgery (i.e. burn wounds that can heal spontaneously but classified as a deep burn wound).

2.6. Synthesis of evidence

Finally, a best evidence synthesis was performed, taking into account the: (1) consistency of results between studies; (2) methodological quality of the studies; and (3) quality of the measurement properties. The overall quality of a measurement property for each technique was displayed as either: strong, moderate, limited, conflicting, or unknown level of evidence (Table 3).

3. Results

3.1. Search

A total of 1664 unique articles were identified with the literature search (see Fig. 1), resulting in 36 articles fulfilling the inclusion criteria, presenting 42 studies (32 human and 10 animal studies). Fourteen measurement techniques were distinguished, classified as (1) laser Doppler techniques; (2)

Table 2 – Quality criteria for acceptable results of studies on measurement properties.

Measurement property	Definition	Adequate (+)	Indeterminate (?)	Inadequate (−)
Reliability	The proportion of the total variance in the measurements, which is due to 'true' differences between patients	ICC or Kappa ≥ 0.7	ICC or Kappa not reported	Criteria for '+' not met
Measurement error	The systematic and random error of a score that is not attributed to true changes in the construct to be measured	SEM or LoA < clinically acceptable values	SEM or LoA not defined	Criteria for '+' not met
Construct validity	The extent to which a particular tool relates to other outcome measures in a manner that is consistent with theoretically derived hypotheses	At least 75% of the results are in accordance with the predefined hypotheses	Only accuracy or differences between relevant groups reported	Criteria for '+' not met

Definitions are in accordance with the COSMIN study [29]. ICC = intraclass correlation coefficient; SEM = standard error of measurement; LoA = limits of agreement.

Table 3 – Levels of evidence for the overall quality of a measurement property.

Level	Criteria
Strong	Consistent findings (adequate or inadequate) in multiple studies of 'good' methodological quality OR in one study of 'excellent' methodological quality
Moderate	Consistent findings (adequate or inadequate) in multiple studies of 'fair' methodological quality OR in one study of 'good' methodological quality
Limited	Findings (adequate or inadequate) in one study of 'fair' methodological quality
Conflicting	Conflicting findings between studies
Unknown	Only studies of 'poor' methodological quality or only indeterminate findings

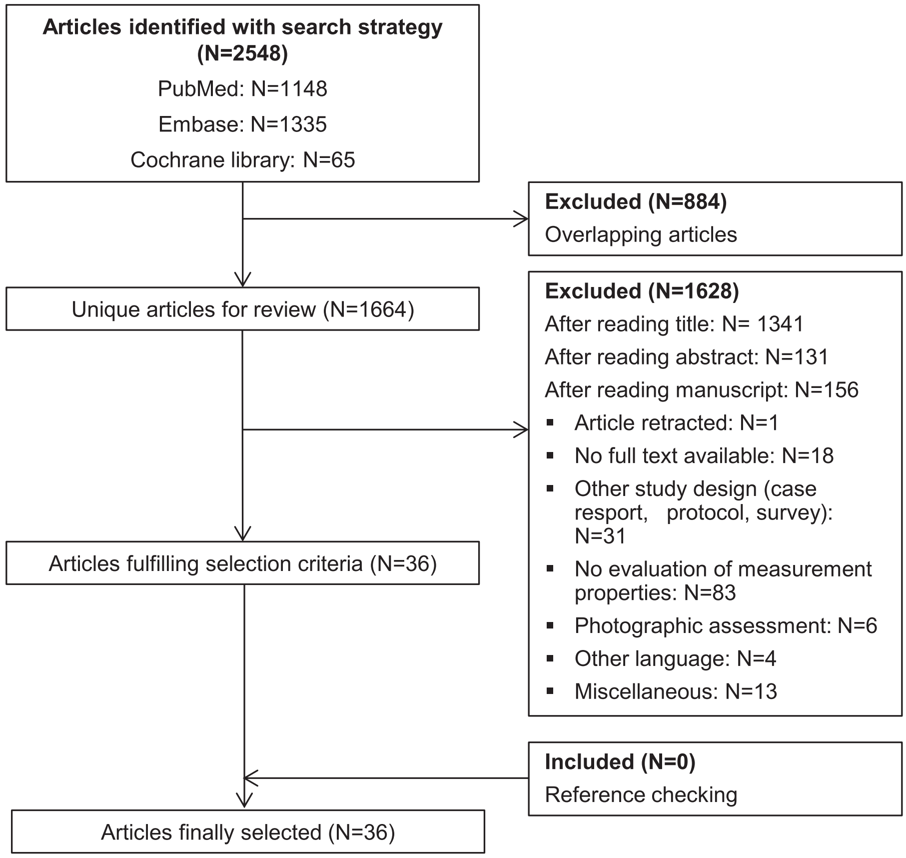

Fig. 1 – Flowchart of literature search and study selection.

thermography or thermal imaging; (3) other measurement techniques. Table 4 shows the included measurement techniques and their characteristics.

3.2. Quality of studies and measurement properties

Details about the methodological quality of included studies and the results on the measurement properties reliability and construct validity are summarized in Tables 5 and 6, respectively. Most studies evaluated construct validity, whereas only two studies investigated the reliability of the measurement technique under study [31,32]. None of the included studies focused on the measurement error. In addition, no studies were found on structural validity, content validity, or internal consistency.

3.3. Synthesis of evidence

Tables 5 and 6 also present the conclusion about the overall quality of respectively the reliability and construct validity, per measurement technique, expressed as the level of evidence.

3.3.1. *Reliability*

Studies that investigated reliability comprised videomicrosopy and dermoscopy as measurement techniques. Both studies evaluated the reliability of the interpretation phase, rather than the reliability of the whole measurement process. We found the quality of the results for the intra- and inter-observer reliability of videomicroscopy conflicting, and thus the level of evidence for this measurement technique was conflicting. For dermoscopy, limited evidence was found for inadequate results.

3.3.2. *Hypotheses testing for construct validity*

Construct validity was evaluated in 42 studies. In category (1) laser Doppler techniques, 18 studies were performed; evaluating laser Doppler flowmetry in the earlier years and, more recently, laser Doppler imaging (LDI). Almost all studies on laser Doppler compared the results to the observed healing time. The most frequently used cut-off point for observed healing was 21 days post-burn, but 14 days post-burn and 12 days post-burn (in children) were also applied. Strong evidence was found for adequate construct validity of LDI.

In category (2) thermography, six studies were performed, evaluating static, active, and active-dynamic thermal imaging. One human study compared thermography to LDI [17]. Another human study compared the results to observed healing in combination with histology (for excised burns), however, no description on the histology results was found [33]. The remaining four studies performed experiments in pigs and used observed healing and/or histology as comparator instrument [16,34–36]. Moderate evidence was found for adequate construct validity of thermography.

Regarding category (3) other measurement techniques, a variety of techniques were evaluated. Many studies in this category performed animal experiments, using rats, mice or pigs, thereby obtaining histological analyses without difficulty [32,37–42]. Ten human studies were performed: three on videomicroscopy [31,43,44], two on spectrophotometry

Table 4 – Characteristics of included measurement techniques.

Measurement technique	Outcome	Mechanism	Completion time	Invasiveness	Interpretation/score calculation	(Dis)advantages	References
Category 1 laser Doppler techniques							
Laser Doppler flowmetry	Perfusion units	780 nm light penetrates the burn wound and scatters back in different wavelengths, based on the velocity of erythrocytes	30 s	Skin contact	Analysis of perfusion units through color coded scale	• Only single point measurements	[56,63,64]
Laser Doppler imaging (LDI)	Perfusion units	640 or 670 nm light penetrates the burn wound and scatters back in different wavelengths, based on the velocity of erythrocytes	4 s–7 min	No skin contact 0.3–0.7 m from burned surface	Analysis of arbitrary perfusion units through color coded scales over a larger surface area than laser Doppler flowmetry	• Long measurement times • High cost • Various PU scales and cut-off values among different laser Doppler devices • Sensitive to patient movements	[17,48–50,55]
Category 2 Thermography or Thermal imaging							
Thermography	Temperature	Infrared light captures the emission of infrared radiation expressed by tissue, which correlates with the remaining perfusion and/or loss of cellular metabolism in the necrotic tissue	Static: immediate Dynamic: several minutes	No skin contact 0.5–0.7 m from burned surface	Temperature difference (ΔT), determined by subtracting the burn wound temperature from the termperature of adjacent or contralateral healthy tissue Static: infrared signals are converted into an image with its respective colors and temperature Dynamic: infrared signals are measured after warming or cooling the tissue with an external source	• Moderate feasibility in dynamic thermography due to cold/warm challenge and requirement for multiple measurements • Large surfaces can be measured • Fast method	[16,17,34–36]
Category 3 Other measurement techniques							
Photoacoustic imaging (PAI)	Chromophores	Chromophores (blood and blood vessels) underneath the burned tissue are selectively excited by light pulses (532 nm fiber laser) and emit thermoelastic waves through adiabatic expansion	Real time: 8–30 frames per second	Skin contact: transducer placed on the tissue surface (burn wound or healthy skin)	The propagation times and amplitudes of the waves give information on respectively the depth and density of chromophores	• Complex analysis	[42]
Spectrophotometric intracutaneous analysis (SIA)	Chromophores	Reflection or transmission of back reflected light from 400 to 1000 nm to obtain calculated maps of chromophores (hemoglobin and melanin)	Immediate	No skin contact 30–50 cm from burned surface	Software converts raw image data into three visible images: high quality photograph, map of perfusion (hemoglobin content) and of pigment (melanin content) that require qualitative interpretation	• Low cost • Easy to use • Difficult interpretation of images	[17]

Table 4 (*continued*)

Measurement technique	Outcome	Mechanism	Completion time	Invasiveness	Interpretation/score calculation	(Dis)advantages	References
Dermoscopy	Presence of microcirculation	Visualization of colors and microstructures of superficial tissue using non-polarized and polarized light, obtained by magnification	Immediate	Skin contact (through plastic sheet) using a handheld dermoscope	Subjective evaluation of the presence or absence of dermoscopic features, which can be described using a graded system of capillary integrity	• Qualitative analysis: subject to observer interpretation • Polarized dermoscopy enables observation of a deeper layer	[32]
Near-infrared spectroscopy (NIRS)	• Oxygen saturation • Hemoglobin content • Water content	Tissue is illuminated by near-infrared rays of three wavelengths: 735, 810, 850 nm, and reflectance is collected by fiber optic probes, providing information on the structural and chemical components of tissue	16–60 s	Skin contact (through plastic sheet)	Regional tissue oxygen saturation (rSO_2) (burn wound rSO_2/normal site rSO_2) as indicator of burn wound depth	• Fast technique	[15]
Videomicroscopy	Presence of microcirculation	Visualization of dermal capillaries by images obtained in the visible light spectrum using a fiber-optic light source and a magnification lens	Immediate	Skin contact (through plastic sheet)	Subjective evaluation of the presence or absence of videomicroscopic features, which can be described using a graded system of capillary integrity	• Qualitative analysis: subject to observer interpretation	[31,44]
Reflectance spectrometry	Range of spectra expressing different chromophores	Reflection or transmission of back reflected light from 400 to 1100 nm to obtain calculated maps of chromophores (hemoglobin and melanin)	?	No skin contact 1 cm from burn wound	Collected spectra are analyzed by the computer using an artificial neural network system to predict the healing time	• Complex analysis • Sensitive to background light and probe distance • Only single point measurements	[45]
Ultrasonography	Skin thickness	Ultrasonographic waves (5 MHz) are transmitted, which correlate with blood flow underneath the burn wound	Immediate	No skin contact 2.5 cm from burn wound	Waves are recorded as individual A-lines, transformed into an image by means of computer processing, which is interpreted by an observer	• Subjective analysis • Low cost	[46]
Spectroscopic optical coherence tomography (OCT)	Tissue scattering spectra	Spectral-domain OCT system with a central wavelength of 850 nm captures light scattered back from tissue	?	?	Processing the OCT signal/raw spectral data by short time fourier transform (STFT) or dual window (DW) method, classifying the tissue as burned or healthy using parameters derived from a power-law or logistic regression classification model	• Complex analysis • Only single point measurements	[39,41]
Dual imaging: Optical coherence tomography	Skin thickness	Light scattered back from tissue is captured to construct a high resolution image	?	Skin contact	The estimated skin thickness is computed from vertical OCT images (i.e. B-scans)	• Penetration depth <1.2 mm • Complex analysis • Sensitive to patient movements	[40]
Pulse speckle imaging (PSI)	Perfusion units	PSI captures the fluctuation in speckle pattern, corresponding	Seconds	No skin contact	Comparison of mean perfusion values	• Large surfaces illuminated at once	

(*continued on next page*)

Table 4 *(continued)*

Measurement technique	Outcome	Mechanism	Completion time	Invasiveness	Interpretation/score calculation	(Dis)advantages	References
		to moving blood cells when illuminated by 785 nm laser light				• Several frames per second	
Spatial frequency domain imaging (SFDI)	Structural and functional parameters	LEDs at 10 wavelengths (400–850 nm) projected off of a spatial light modulator, collected with a near-infrared camera	5 ms integration times	No skin contact	Quantitative spatial maps of tissue optical properties and biochemical composition such as tissue chromophores (total hemoglobin and oxygen saturation)	• Capable at each wavelength • Relatively large surfaces can be measured $(30 \times 20\,cm^2)$	[38]
Fiber-optic confocal imaging (FOCI)	Thickness	Detection of burn-associated autofluorescence (due to denatured collagen), which is laser-induced (excitation: 488 nm, detection: >505 nm)	?	Skin contact	Thickness of the burn-associated autofluorescent layer estimated from the distance between the most superficial and deepest detectable fluorescent XY images	• Multiple factors confound interpretation of data • Only single point measurements	[37]
Multispectral imaging	Tissue absorption spectra	Reflectance of red/green/ near-infrared light using an optical probe and electronic control box	?	No skin contact	Data are read directly of the dials of the device and can be plotted in a grid to characterize the burn wound	• Sensitive to reflection from the skin	[8,47]

PU: perfusion units. ?: no information available.

Table 5 – Reliability: methodological quality, quality of results and overall level of evidence per measurement technique.

Measurement technique: tool	First author country, year	N	COSMIN methodological quality	Time interval	Results	Quality of results (rating)
Videomicroscopy: Bscan-pro [31]	Mihara Japan, 2012	44 pt 56 bw	FAIR	Analyses were performed retrospectively on a single measurement/photo (interpretation phase)	Intra-observer (unweighted) κ: 0.72, 0.73 Inter-observer (unweighted) κ: 0.64	+ –
Level of evidence videomicroscopy: UNKNOWN						
Dermoscopy: Ondeko Dermoscope Epilite x8 [32]	Mihara Japan, 2015	30 bw	FAIR	Analyses were performed retrospectively on a single measurement/photo (interpretation phase)	Inter-observer (unweighted) κ: 0.39, 0.40, 0.25	–
Level of evidence dermoscopy: LIMITED for inadequate results						

pt = patients; bw = burn wounds; NA = not assessed; ICC = intraclass correlation coefficient; SEM = standard error of measurement; LoA = limits of agreement.

[17,45],one on dermoscopy [32], one on near-infrared spectroscopy [15], one on ultrasonography [46], and two on multispectral imaging [8,47]. In this category, the outcomes of the measurement techniques were mainly compared to observed healing and in two studies to LDI. Moderate evidence was found for adequate construct validity of videomicroscopy and spatial frequency domain imaging (SFDI). Limited evidence was found for adequate construct validity of dermoscopy, near-infrared spectroscopy (NIRS), reflectance spectrometry, and dual imaging (pulse-speckle imaging (PSI) and optical coherence tomography (OCT)). Unknown evidence was found for fiber-optic confocal imaging (FOCI), multispectral imaging, spectroscopic OCT, ultrasonography, spectrophotometric intracutaneous analysis (SIA), and photoacoustic imaging (PAI).

Furthermore, we found one explanatory article on LDI [48]. The article by Pape et al. is not described in Tables 5 or 6 as it did not provide information on the measurement properties reliability, measurement error or construct validity. Therefore, the article is not evaluated in the same way as the other included articles. Nevertheless, the study appeared in our systematic search and provides important information on the LDI color palette, based on healing times of a series of burn wounds in adults and children. The color palette considers the following boundaries: healing within 14 days is red (PU > 600), healing between 14 and 21 days is yellow (PU 261-440), and healing beyond 21 days is blue (PU < 200). The intermediate categories are formed by pink (PU 440-600); healing at about 14 days, and green (PU 201-260); healing at about 21 days. The category blue is further subdivided into dark blue (PU 0-140) for burn wounds that have 96% probability for no healing by day 21, and light blue (PU 141-200) with 74.4% probability for no healing by day 21. Monstrey et al. subsequently validated the color palette [49], showing an overall accuracy of 96.3%.

4. Discussion

The purpose of this review was to critically appraise, compare and summarize the quality of techniques that aim to assess burn wound depth or healing potential. Ultimately, we aimed to provide a recommendation on the most suitable technique. If only high quality outcome measurement techniques will be used in the future, the quality of clinical studies using these techniques will consequently improve. To our knowledge, this is the first systematic review on techniques for burn wound assessment that took the methodological quality of included studies into account.

Strong evidence was found for adequate results on construct validity for LDI. Therefore, LDI seems the most favorable measurement technique for burn wound assessment, thereby determining burn wound healing potential. The methodological quality of studies that investigated construct validity of LDI varied considerably; nevertheless, four studies were scored as ‘good’ and one as ‘excellent’, in which adequate descriptions of the comparator instrument (e.g. observed healing), including its measurement properties, were given, and appropriate statistical analyses were used. The comparability of study results in the laser Doppler category is challenging, as diverse tools were evaluated, using differing cut-off perfusion values. Ten studies used the LDI Burn Imager [2,17,44,49–52] or Laser Doppler Line Scanner (LDLS) [53,54], both from Moor Instruments, Axminster, United Kingdom. Furthermore, other laser Doppler tools were evaluated such as the Periscan PIM 3 [55] or PF 4001 [56] (Perimed AB, Stockholm, Sweden). These systems maintain other perfusion cut-off values compared to the Moor systems to determine whether a burn wound will heal. Nevertheless, the results of all LDI tools, developed by different companies, showed consistent high sensitivity and specificity values (varying between 74% and 100%), if the scan was performed >48 h post-burn. For the purpose of clarity and because of these consistent results, we decided to present one level of evidence for LDI as technique, rather than for every LDI tool separately. Nevertheless, the evidence per measurement tool can be obtained from Table 6. Furthermore, we found one explanatory article on LDI [48]. To pursue uniformity, we recommend using the color palette of Pape et al. to make predictions about the healing potential of burn wounds assessed by LDI. One of the advantages of LDI is that the technique has been developed in such a way that a largepart of the analysis and image interpretation is

Table 6 – Construct validity: methodological quality, quality of results and overall level of evidence per measurement technique.

Measurement technique: tool	First author country, year	N	COSMIN methodological quality	Comparator instrument (i.e. LDI, observed healing, histology or clinical evaluation)	Results	Quality of results (rating)	Comments
Category 1 laser Doppler techniques							
LDI: moorLDI2-BI [17]	Burke-Smith UK, 2015	18 pt 24 bw	GOOD	Observed healing Group A: healing <14 days Group B: healing 15–21 days Group C: healing >21 days	Accuracy 74% (17/23 burn areas) Group A vs. Group B: $p<0.001$ Group B vs. Group C: $p>0.05$	?	Reference for measurement property of observed healing
LDLS: moorLDLS-BI [53]	Hoeksema Belgium, 2014	204 pt (adults and children) 321 bw	FAIR	Observed healing	Sens 91.9%, Spec 96%, PPV 91.9%, NPV 96%, Accuracy 94.2%	+	Scarce information on observed healing
		198 sites	EXCELLENT	LDI: moorLDI2-BI	Sens 92.3%, Spec 96.1, PPV 92.3, NPV 96.1%, Accuracy 94.4%	+	
					Correlation between LDLS and LDI: $R^2=0.9095$ (N = 77 bw)	+	
LDLS: moorLDLS-BI [54]	Holland Australia, 2014	49 pt (children) 59 bw 90 scans	FAIR	Observed healing <14, 14–21 days (or surgery), >21 days	<14 days: Sens 98%, Spec 79%, Accuracy 96%	+	Scarce information on observed healing
					14–21 days: Sens 70%, Spec 95%, Accuracy 94%	+	
					>21 days: Sens 92%, Spec 97%, Accuracy 95%	+	
LDI: moorLDI2-BI					<14 days: Sens 97%, Spec 72%, Accuracy 95%	+	
					14–21 days: Sens 71%, Spec 96%, Accuracy 94%	+	
					>21 days: Sens 70%, Spec 100%, Accuracy 95%	+	
LDI: moor LDI imagers [49]	Monstrey Belgium, UK, USA, 2011	433 bw	FAIR	Observed healing <14, 14–21 or >21 days	Sens 94.5%, Spec 97.2%, PPV 94.5%, NPV 97.2%, Accuracy 96.3% Only parameter that influences the accuracy is gender: 1.1%	+	Scarce information on observed healing
Laser Doppler flowmetry: O2C [20]	Merz Germany, 2010	28 pt 173 bw	FAIR	Observed healing <1, <2 or <3 weeks, or surgery	Sens 80.6%, Spec 88.2%, PPV 93.1%, NPV 69.8%, Accuracy 83.2%, AUC 0.94 (N = 101 bw)	+	Cut-off value: flow of 100 arbitrary units Scarce information on observed healing
LDI: moorLDI2-BI [50]	Nguyen	924 pt	FAIR	Observed healing <21 or >21 days	>48 h post-burn: Sens 76.2%, Spec 80.8%, AUC 0.81	+	Scarce information on observed healing and LDI interpretation
	Australia, 2010	638 bw			<48 h post-burn: Sens 78%, Spec 74%, AUC 0.78	+	

Table 6 *(continued)*

Measurement technique: tool	First author country, year	N	COSMIN methodological quality	Comparator instrument (i.e. LDI, observed healing, histology or clinical evaluation)	Results	Quality of results (rating)	Comments
LDI: Periscan PIM III [55]	Cho Republic of Korea, 2009	103 pt (<15 years) 181 bw	GOOD	Observed healing <14 or >14 days (only conservatively treated bw were included)	Sens 80.6%, Spec 76.9%, AUC 0.84	+	Cut-off value LDI: 250 perfusion units Two observers assessed observed healing
LDI: moorLDI2-BI [2]	Hoeksema Belgium, 2009	40 pt 40 bw	GOOD	Observed healing <21 or >21 days (biopsy of surgically treated wounds <21 days)	Accuracy (%): Day 0: 54 Day 1: 79.5 Day 3: 95 Day 5: 97 Day 8: 100 PPV, NPV (%): Day 0: 42.9, 64.7 Day 1: 66.7, 90.5 Day 3: 87.5, 100 Day 5: 100, 95.7 Day 8: 100, 100	 ? ? ? ? ? – – + + +	Cut-off value LDI: <220 perfusion units to predict non-healing at day 21 Two observers assessed observed healing No information on histological analysis
LDI: Moor LDI scanner [44]	McGill UK, 2007	20 pt 27 bw	FAIR	Observed healing <21 or >21 days, or surgery	Inversed correlation Kendall's Tau r = −0.697 (p < 0.001)	–	Scarce information on observed healing
LDI: Moor LDI version 2 [51]	La Hei Australia, 2006	31 pt 50 scans	FAIR	Observed healing <14 or >14 days	Sens 97%, Spec 100%	+	Adequate information on healing, but measurement properties unknown
LDI: Moor LDI-VR [62]	Jeng USA & UK, 2003	23 pt 41 bw	POOR	Histology Clinical evaluation	Accuracy 100% (7/7 wounds) LDI and surgeon agreed in 56% of cases, p = 0.013	?	No information on histology and methodological flaws on LDI interpretation
Laser Doppler Flowmeter: PF 4001 [56]	Mileski USA, 2003	51 pt 153 bw	POOR	Observed healing <21 days or >21 days/surgery	Day 0: Sens 94%, Spec 54%, PPV 62%, NPV 92% Day 1: Sens 94%, Spec 51%, PPV 60%, NPV 91% Day 2: Sens 91%, Spec 66%, PPV 76%, NPV 86%	– – +	Cut-off value: <80 perfusion units Statistical flaws

(continued on next page)

Table 6 (continued)

Measurement technique: tool	First author country, year	N	COSMIN methodological quality	Comparator instrument (i.e. LDI, observed healing, histology or clinical evaluation)	Results	Quality of results (rating)	Comments
					Day 3: Sens 85%, Spec 78%, PPV 76%, NPV 86%	+	
					Serial measurements: Sens 80%, Spec 88%, PPV 81%, NPV 84%	+	
					*Day 0: Sens 63%, Spec 81%, PPV 86%, NPV 54%	+	*Calculated by Table 3 of the original study
					*Day 1: Sens 63%, Spec 87%, PPV 92%, NPV 51%	+	
					*Day 2: Sens 66%, Spec 84%, PPV 84%, NPV 66%	+	
					*Day 3: Sens 77%, Spec 81%, PPV 81%, NPV 78%	+	
					*Serial measurements: Sens 81%, Spec 77%, PPV 67%, NPV 88%	+	
LDI: PIM II [65]	Riordan USA, 2003	22 pt 35 bw	GOOD	Histology: hematoxylin and eosin, and vimentin staining	Sens 95%, Spec 94% Correlation between LDPI and burn depth (μm): r = −0.84 (p < 0.01)	+	Adequate information on histological analysis including reference
LDI: Moor LDI system [52]	Holland Australia, 2002	57 pt (children) 57 bw	FAIR	Observed healing <12 or >12 days	Sens 90%, Spec 96%	+	Scarce information on observed healing
					PPV 96%, NPV 90%	+	Calculated by Table 3 of the original study
Laser Doppler: Periflux 4000 [60]	Yeong Taiwan, 1996	152 bw	FAIR	Observed healing <14 days or >14 days/surgery	Accuracy 94%, Sens 97%, Spec 92%	+	Adequate information on observed healing, but measurement properties unknown
Laser Doppler Flowmeter [63]	Atiles USA, 1995	22 pt 57 bw	FAIR	Observed healing <21 or >21 days	At day 2 post-burn as indicative for healing >21 days: Sens 88%, Spec 100%, PPV 100%, NPV 92%	+	Cut-off value: <40 perfusion units Scarce information on observed healing
Laser Doppler Flowmeter [64]	Atiles USA, 1995	21 pt 86 bw	FAIR	Observed healing <21 or >21 days	At day 3 post-burn as indicative for healing >21 days: Sens 46%, Spec 100%, PPV 100%, NPV 85%	+	Cut-off value: <40 perfusion units Scarce information on observed healing

Table 6 *(continued)*							
Measurement technique: tool	First author country, year	N	COSMIN methodological quality	Comparator instrument (i.e. LDI, observed healing, histology or clinical evaluation)	Results	Quality of results (rating)	Comments
Level of evidence LDI: STRONG for adequate results							
Category 2 thermography or thermal imaging							
Infrared thermography (IRT): FLIR E60 [17]	Burke-Smith UK, 2015	18 pt 24 burn regions	GOOD	LDI	Correlation between IRT and LDI: R=0.73 Group A vs. Group B: p < 0.01 Group B vs. Group C: p > 0.05 Group A: healing <14 days Group B: healing 15–21 days Group C: healing >21 days	+	Adequate information on LDI
Thermography: FLIR T300 [33]	Singer USA, 2015	24 pt 39 bw	FAIR	Observed healing ≤21 or >21 days, and histology (hematoxylin and eosin staining) for excised burns	Sens 87%, Spec 87.5%, PPV 90.9%, NPV 82.4%, Accuracy 87.2%, AUC 0.83 (95% CI 0.69–0.97) Optimal cut-off point 0.1 °C, which was rounded to 0 °C	+	No report of histology results, scarce information on observed healing
					Sens (14/16) 87.5%, Spec (20/23) 87%, PPV (14/17) 82.4%, NPV (20/22) 90.9%, Accuracy (34/39) 87.2%	+	*Calculated by Table 1 of the original study
Active dynamic thermography [35]	Renkielska Poland, 2014	? pigs 65 bw	POOR	Histology: <60% or >60% of dermis thickness	<60%: time constant=49.1 ±23.0 s >60%: time constant=86.3 ±32.3 s Difference p < 0.05	?	Results are in contrast with article from 2006
					Accuracy 83.9%, Sens 77.8%, Spec 89.7% (not clear whether these results are based on comparison with histology or healing)	+	
Thermal infrared imaging: AGEMA THV-900 SW/TE [36]	Ruminksy Poland, 2007	3 pigs 23 bw	FAIR	Histology: percentage of dermis thickness at the measurement site (dtms); <60% (will heal) or >60% (will not heal in 3 weeks)	Static imaging: AUC 0.94, Sens 80%, Spec 100% (p < 0.001 ANOVA)	+	Statistical flaws
					Active imaging: AUC 0.75 (p=0.07 ANOVA)	+	
					Active-dynamic imaging: AUC 1.0 (p=0.003 ANOVA)	+	

(continued on next page)

Table 6 (continued)

Measurement technique: tool	First author country, year	N	COSMIN methodological quality	Comparator instrument (i.e. LDI, observed healing, histology or clinical evaluation)	Results	Quality of results (rating)	Comments
					*Correlation histology and static imaging $R=-0.67$, $p=0.001$ *Correlation histology and active imaging $R=0.29$, $p=0.175$ *Correlation histology and active-dynamic imaging $R=-0.37$, $p=0.081$	– – –	*Calculated by Table 2 of the original study
Active dynamic thermography [34]	Renkielska Poland, 2006	3 pigs 23bw	FAIR	Observed healing <21 or >21 days Histology: <60% or >60% of dermis thickness	Accuracy, Sens and Spec: 100% <60%: time constant=12.08 ±1.94s >60%: time constant=9.07 ±0.68s Difference $p<0.05$	+ ?	Statistical flaws: wrong definition of sensitivity in article
Static thermography [16]	Renkielska Poland, 2005	11 pigs 64bw	POOR	Observed healing <21 or >21 days Histology: <60% or >60% of dermis thickness Clinical evaluation: II°a, II°b, III°	Accuracy 93.8%, Sens 97.7%, Spec 85.5% <60%: 0.88°C ±0.62 >60%: −0.48°C ±0.65 Difference $p<0.001$ II°a: 0.96°C ±0.54; II°b: 0.77°C ±0.73; III°: −0.44°C ±0.70 Difference between II°a and III° $p<0.001$ Difference between II°b and III° $p<0.01$	+ ? ?	Statistical flaws: wrong definition of sensitivity in article
Level of evidence thermography: MODERATE for adequate results							
Category 3 Other measurement techniques							
Photoacoustic imaging (PAI) system [42]	Ida Japan, 2016	4rats/bw temperature 24 rats in total	POOR	Histology: viable blood vessels Burn induction temperature: 70, 78, 83, 88, 93, 98°C	Correlation $R^2=0.83$, $R=0.91$, $p=0.00$ Correlation $R^2=0.95$, $R=0.97$, $p=0.00$	+ +	No information on measurement properties histology
Level of evidence PAI: UNKNOWN							
Spectrophotometric intracutaneous analysis (SIA): Scanoskin^{TM} [17]	Burke-Smith UK, 2015	18pt 24 burn regions	FAIR	Observed healing Group A: healing <14 days Group B: healing 15–21days Group C: healing >21 days	Qualitative analyses: Group A and B darker red, representing increased hemoglobin density due to increased perfusion. Group A showed	?	Reference for measurement property of observed healing, but statistical methods not optimal

Table 6 *(continued)*

Measurement technique: tool	First author country, year	N	COSMIN methodological quality	Comparator instrument (i.e. LDI, observed healing, histology or clinical evaluation)	Results	Quality of results (rating)	Comments
					more pronounced increase in perfusion compared to group B. Group A and B both showed reduced pigment. Group C showed almost white, representing reduced hemoglobin density due to reduced perfusion. Mixed pigmentation pattern.		
Level of evidence SIA: UNKNOWN							
Dermoscopy: Ondeko Dermoscope Epilite x8 [32]	Mihara Japan, 2015	? pt 30 bw	FAIR	Observed healing <21 days or >21 days Clinical assessment	Accuracy 96.7%, Sens 100%, Spec 94.4% Accuracy 76.7%, Sens 85.7%, Spec 73.9%	+ +	Scarce information on observed healing
Level of evidence dermoscopy: LIMITED for adequate results							
Near-infrared spectroscopy (NIRS): NIRO-200NX [15]	Seki Japan, 2014	14 pt 50 bw	FAIR	LDI (measurements <48 h post-burn)	Spearman correlation: r = 0.755, p < 0.001 After standardization of O2 values (to exclude influence of systemic circulation): r = 0.678, p < 0.001	+ –	Scarce information on LDI interpretation of perfusion units
Level of evidence NIRS: CONFLICTING							
Videomicroscopy: Bscan-pro [31]	Mihara Japan, 2012	44 pt 56 bw	FAIR	Observed healing <21 days or >21 days	Sens 81.8%, Spec 100%, PPV 100%, NPV 89.5%	+	Scarce information on observed healing
Videomicroscopy: B-scan-pro [43]	Mihara Japan, 2011	41 pt 44 bw	FAIR	Observed healing <21 days or >21 days	Overall: Sens 81.8%, Spec 87.8%, PPV 69.2%, NPV 93.5%, Accuracy 86.4% Groups: <24 h, 24–62 h and >62 h post-burn; in between-group comparison revealed a significantly lower accuracy in the <24 h group	+	Scarce information on observed healing
Videomicroscopy [44]	McGill UK, 2007	20 pt 27 bw	FAIR	Observed healing <21 or >21 days, or surgery	Correlation Kendall's Tau r = 0.751 (p < 0.001)	+	Scarce information on observed healing

(continued on next page)

Table 6 (*continued*)

Measurement technique: tool	First author country, year	N	COSMIN methodological quality	Comparator instrument (i.e. LDI, observed healing, histology or clinical evaluation)	Results	Quality of results (rating)	Comments
				LDI	Inversed correlation Kendall's Tau $r = -0.725$ ($p < 0.001$)	?	
Level of evidence videomicroscopy: MODERATE for adequate results							
Reflectance spectrometry: SD-2000 [45]	Yeong Taiwan, 2005	39 pt 47 bw	FAIR	Observed healing <14 or >14 days	Sens 75%, Spec 97%, Accuracy 86%	+	Scarce information on observed healing
Level of evidence reflectance spectrometry: LIMITED for adequate results							
Ultrasonography [46]	Iraniha USA, 2000	15 pt 78 bw 42 controls	POOR	Observed healing <21 or >21 days (or surgery)	Accuracy 96%, Sens 100%, Spec 92%	+	Statistical flaws, positive and negative exchanged
Level of evidence ultrasonography: UNKNOWN							
Spectroscopic optical coherence tomography (OCT) [41]	Zhao USA, 2015	4 mice?	POOR	Histology: epidermal, superficial partial, deep partial, full thickness; not specified	Power law: AUC 0.70, accuracy 65%	+	Processing: short time Fourier transform (STFT) method
					Logistic regression: AUC 0.84, accuracy 76%	+	
					Negative correlation: −0.9998 between burn depth and normalized power law exponent ($p < 0.001$)	+	No information on histology
Spectroscopic optical coherence tomography (OCT) [39]	Maher USA, 2014	mice 6 bw 4 healthy sites	POOR	Histology: qualitative description after hematoxylin-eosin staining	DW: AUC 0.83, 0.87 Accuracy 76%, 80%	+	Processing OCT data: short time
					STFT: AUC 0.95, 0.97 Accuracy 90%, 91%	+	Fourier transform (STFT) and dual window (DW) method No information on measurement properties histology
Level of evidence spectroscopic OCT: UNKNOWN							
Dual-imaging: Optical coherence tomography (OCT) and pulse speckle imaging (PSI) [40]	Ganapathy USA, 2014	3 pigs 68 bw	FAIR	Histology: Superficial (S) <750 μm Partial (P) 750–1750 μm Full (F) >1750 μm	Dual-imaging: AUC: 0.94 (S), 0.79 (P), 0.87 (F): $p < 0.001$ PSI: averaged AUC 0.78 OCT: averaged AUC: 0.62	+	No information on measurement properties histology

Table 6 (continued)

Measurement technique: tool	First author country, year	N	COSMIN methodological quality	Comparator instrument (i.e. LDI, observed healing, histology or clinical evaluation)	Results	Quality of results (rating)	Comments
					Dual-imaging averaged AUC: 0.85		
Level of evidence dual-imaging: LIMITED for adequate results							
Spatial frequency domain imaging (SFDI) [38]	Mazhar USA, 2014	4 pigs 48 bw	GOOD	Histology: vimentin immunostain and Masson's trichrome	Correlation SFDI single wavelength (658 nm) and histology $r^2 = 0.94$ for all burn severities for time points >24 h Correlation tissue scattering parameters and histology (vimentin): $r^2 > 0.89$	+	Adequate information on histology
Level of evidence SFDI: MODERATE for adequate results							
Fiber-optic confocal imaging (FOCI) [37]	Vo Australia, 2001	? mice	POOR	Histology: depth of collagen damage (μm) and estimated thickness of autofluorescent layer by FOCI (μm)	Correlation r = 0.78	+	No information on measurement properties histology
Level of evidence FOCI: UNKNOWN							
Multispectral imaging: Burn Depth Indicator [47]	Afromowitz USA, 1988	32 pt 112 bw	FAIR	Observed healing <21 or >21 days	All burns: accuracy 88% Intermediate burns (n = 55): accuracy 84%	?	Scarce information on observed healing
Multispectral imaging: Burn Depth Indicator [8]	Heimbach USA, 1984	? pt ? bw 569 points	POOR	Observed healing <21 or >21 days	Determining healing in 21 days: overall accuracy 76%, bw predicted to heal 77% and bw predicted not to heal 74% Determining healing in 31 days: predicted to heal 91% and predicted not to heal 52%	?	Methodological flaws
					*Accuracy 84%, Sens 84%, Spec 84%	+	*Calculated by Table 1 of the original study
Level of evidence multispectral imaging: UNKNOWN							

LDI = laser Doppler imaging; LDLS = laser Doppler line scanner; pt = patients; bw = burn wounds; AUC = area under the receiver operating curve; Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value; *these results were not directly presented in the included manuscript, but retrieved or calculated from raw data.

standardized, and therefore the observer can minimally affect it. Only the area of interest has to be traced manually, but thereafter the computer performs automatic analysis of the mean perfusion value. However, until now, no studies on the reliability of this specific phase in the performance of LDI were found, whilst it is known from previous research that manual tracing of the area of interest can certainly influence the reliability [57]. Therefore, we would like to stress the need to examine these sources of variation in further research. Likewise, the analysis of the output of techniques such as videomicrosopy, dermoscopy, ultrasonography, and spectrophotometry, is not standardized at all. In these techniques, interpretation of the obtained data is required using a scoring algorithm [43,44], which is possibly accompanied by a high-degree of variation between observers or over time. It is essential to take into account that these measurement techniques comprise a large subjective component and therefore it is important to evaluate the intra- and interobserver reliability. Indeed, it was surprising that we found only two studies focusing on the reliability of the measurement technique under study [31,32]. It is recommended that in future studies, multiple measurements are performed in stable patients by at least two observers [58]. Additionally, as reliability parameters are highly dependent on the variation in the studied sample (i.e. heterogeneity), it is of importance to analyze a sample that reflects the population in clinical practice. Moreover, to be able to monitor an individual patient over time in clinical practice, agreement parameters such as the measurement error and the smallest detectable change should be determined. These parameters are expressed on the actual scale of measurement and can therefore easily be interpreted by clinicians [59].

One of the reasons that many studies received a poor score for the assessment of construct validity is because the authors failed to provide adequate information on the comparator instrument. Most of the time there was no specification, sometimes ‘observed healing’ was defined as complete epithelialization of the burn wound, and in another study it was defined as when no dressings were needed [60]. Additionally, the measurement properties of ‘observed healing’ were often not reported and it was unclear whether one or two observers performed this type of assessment. Consequently, the use of ‘observed healing’ as comparator instrument was unwarranted in many studies. In future research, it is recommended to define ‘burn wound healed’ as the post-burn day on which $\geq$95% of the original wound surface area is epithelialized, which has to be performed preferably by an experienced observer (i.e. more than 10 years practice in burns) or alternatively by two less experienced observers [61]. The statement that healing is obtained when no bandages are needed is not recommended, as this method is susceptible to different interpretations.

Another comparator instrument for which adequate information was often lacking was histology. Several studies did not precisely describe the specific content and measurement properties of their type of histological analysis [37,39,41,42,62]. It is difficult to evaluate the measurement technique of interest in an adequate way when the measurement properties of the comparator instrument are not known or uncertain. Moreover, the problem with histology is that many different characteristics of a biopsy sample can be analyzed, such as the level of viable versus thrombosed blood vessels (i.e. microvascular injury), the presence of collagen changes or level of collagen damage, and the extent of tissue damage (i.e. percentage of dermis thickness destroyed). The selected characteristic for analysis appears to vary between burn centers and may depend on the experience of the local pathologist or analyst. As a result, each characteristic provides specific information on the depth of a burn wound and may therefore be assigned as a single subconstruct. Therefore, choosing histology as comparator instrument can be relevant, but to draw conclusions about validity, it is necessary to describe the precise subconstruct that is assessed.

Another reason why some studies received a fair or poor score is because we noticed that the given results were not calculated in an adequate manner. In one study on thermography, the raw data were presented in the manuscript, allowing us to recalculate the presented sensitivity, specificity, PPV, and NPV, leading to different results as presented in the original article (see Table 6) [47]. In another study, it was possible to calculate the correlations between histology and static, active, and active-dynamic imaging, besides the AUCs that were already established, resulting in an altered conclusion (see Table 6) [36]. The minimum requirement is to calculate the sensitivity and specificity for dichotomous scores, and the correlation or AUC for continuous scores. However, the difficulty with sensitivity and specificity within the assessment of burn wounds is: what is defined as a ‘positive condition’ and as a ‘positive test’? This can easily be interchanged, as evidenced by some included studies [16,33,46,56]. It is considered here that a positive condition is generally stated as ‘having the disease’. When this is translated into burn wound depth or healing potential, we conclude that burn wounds that do not heal or have a healing potential $>$21 days (or $>$14 days if this is the preferred cut-off point), should be defined as a positive condition, thereby reflecting the situation with the worst outcome. On the contrary, burn wounds with a healing potential $<$21 days (or $<$14 days) are defined as a negative condition. Furthermore, a positive test confirms the condition. For example in LDI, a positive test can be reflected by measuring less than 200 PU, or in thermography, it can be reflected by a low ΔT (e.g. $-2.0^{\circ}C$). As a result, the sensitivity (i.e. the true positive rate) is the proportion of positives that are correctly identified as such (i.e. the percentage of patients who are correctly identified as having the condition/no healed burns by the measurement technique). The specificity (i.e. true negative rate) is the proportion of negatives that are correctly identified as such (i.e. the percentage of patients who are correctly identified by the measurement technique as not having the condition/healed burns). To be able to adequately compare clinical studies, it is of great importance that the interpretation and use of these terms becomes uniform.

With regard to the construct to be measured in burn wound assessment, we would like to note the following. Depth and healing potential are frequently used interchangeably, whilst it seems that both terms comprise a different construct. Healing potential reflects the potential for re-epithelialization, which can be achieved by surviving keratinocytes or epidermal stem cells from the sweat ducts, hair follicles, and sebaceous glands (i.e. appendages in the dermis). As these appendages have a wealthy blood supply, it appears that blood flow is an adequate indicator to determine healing potential. On the other hand, burn wound depth seems more a reflection of the extent of tissue damage or necrosis of the skin or underlying tissues (e.g. the subcutis). For future research, it is important that studies provide an appropriate description or definition of which construct exactly is aimed to measure. When hypotheses are formulated about expected direction and magnitude of relations with other instruments, these comparator instruments should also be appropriately described. For example, if the comparator instrument is histological analysis, which intends to assess burn wound depth, it should be described which subconstruct exactly it aims to measure. Accordingly, the relation between the measured constructs can be ascertained.

Regarding future developments, it is considered that it might be useful to apply two measurement techniques at the same time to acquire more information than when using a single technique. One study reported on a dual imaging system consisting of PSI and OCT, measuring the perfusion and birefringence property of collagen, respectively. In this way, by measuring two subconstructs, the validity of burn wound assessment may further improve. The authors showed improved results (AUC 0.85) compared to the individual modalities (PSI: AUC 0.78; OCT: AUC 0.62) [40]. It would be interesting to progressively find out if fusion of information across different measurement techniques can improve the sensitivity and specificity of assessing burn wound depth or healing potential.

5. Conclusions

Based on the quality of included studies evaluating measurement techniques for the assessment of burn wounds and the available evidence, it appears that LDI is the most favorable technique; thereby assessing burn wound healing potential. However, we recommend testing the reliability of LDI to a greater extent. Furthermore, it would be appropriate to select one or two measurement techniques that have the potential to be used in the future, and to further develop and evaluate these techniques. This may depend on the access and expertise in a specific burn center. We would like to underpin the importance of extensive testing of a technique's measurement properties by using adequate outcome parameters and determining the measurement error. In addition, it is underscored that new studies need to develop uniform construct definitions or at least provide a precise description of which construct exactly is aimed to measure. This could lead to improved quality of studies on measurement techniques for burn wound assessment and finally improve burn care.

Conflict of interest

The authors state that they have no financial or other conflict of interest in the publication of this article.

Acknowledgements

This project was supported by a grant from the Dutch Burns Foundation (grant number 13.107). We would like to thank Ralph de Vries, librarian of the VU University Medical Center, for his help concerning the literature search.

Appendices A and B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.burns.2018.05.015.

REFERENCES

[1] Singh V, Devgan L, Bhat S, Milner SM. The pathogenesis of burn wound conversion. Ann Plast Surg 2007;59:109–15.
[2] Hoeksema H, Van de Sijpe K, Tondu T, Hamdi M, Van Landuyt K, Blondeel P, et al. Accuracy of early burn depth assessment by laser Doppler imaging on different days post burn. Burns 2009;35:36–45.
[3] Herndon DN. Total burn care. third ed. Saunders Elsevier; 2007.
[4] Heimbach D, Engrav L, Grube B, Marvin J. Burn depth: a review. World J Surg 1992;16:10–5.
[5] Monstrey S, Hoeksema H, Verbelen J, Pirayesh A, Blondeel P. Assessment of burn depth and burn wound healing potential. Burns 2008;34:761–9.
[6] Devgan L, Bhat S, Aylward S, Spence RJ. Modalities for the assessment of burn wound depth. J Burns Wounds 2006;5:e2.
[7] Godina M, Derganc M, Brcic A. The reliability of clinical assessment of the depth of burns. Burns 1977;4:92–6.
[8] Heimbach DM, Afromowitz MA, Engrav LH, Marvin JA, Perry B. Burn depth estimation—man or machine. J Trauma 1984;24:373–8.
[9] Jackson DM. The diagnosis of the depth of burning. Br J Surg 1953;40:588–96.
[10] Watts AM, Tyler MP, Perry ME, Roberts AH, McGrouther DA. Burn depth and its histological measurement. Burns 2001;27:154–60.
[11] Meyerholz DK, Piester TL, Sokolich JC, Zamba GK, Light TD. Morphological parameters for assessment of burn severity in an acute burn injury rat model. Int J Exp Pathol 2009;90:26–33.
[12] Kahn AM, McCrady VL, Rosen VJ. Burn wound biopsy. Multiple uses in patient management. Scand J Plast Reconstr Surg 1979;13:53–6.
[13] Chvapil M, Speer DP, Owen JA, Chvapil TA. Identification of the depth of burn injury by collagen stainability. Plast Reconstr Surg 1984;73:438–41.
[14] Robson MC, Kucan JO, Paik KI, Eriksson E. Prevention of dermal ischemia after thermal injury. Arch Surg 1978;113:621–5.
[15] Seki T, Fujioka M, Fukushima H, Matsumori H, Maegawa N, Norimoto K, et al. Regional tissue oxygen saturation measured by near-infrared spectroscopy to assess the depth of burn injuries. Int J Burns Trauma 2014;4:40–4.
[16] Renkielska A, Nowakowski A, Kaczmarek M, Dobke MK, Grudzinski J, Karmolinski A, et al. Static thermography revisited—an adjunct method for determining the depth of the burn injury. Burns 2005;31:768–75.
[17] Burke-Smith A, Collier J, Jones I. A comparison of non-invasive imaging modalities: infrared thermography, spectrophotometric intracutaneous analysis and laser Doppler imaging for the assessment of adult burns. Burns 2015;41:1695–707.
[18] Niazi ZB, Essex TJ, Papini R, Scott D, McLean NR, Black MJ. New laser Doppler scanner, a valuable adjunct in burn depth assessment. Burns 1993;19:485–9.
[19] Droog EJ, Steenbergen W, Sjoberg F. Measurement of depth of burns by laser Doppler perfusion imaging. Burns 2001;27:561–8.

[20] Merz KM, Pfau M, Blumenstock G, Tenenhaus M, Schaller HE, Rennekampff HO. Cutaneous microcirculatory assessment of the burn wound is associated with depth of injury and predicts healing time. Burns 2010;36:477–82.
[21] Elamin SE, Dickson JK, Mackie IP. Is Laser Doppler imaging (LDI) a measure of burn depth? Burns 2015;41:413.
[22] Monstrey S, Hoeksema H, Verbelen J, Pirayesh A, Blondeel P. Assessment of burn depth and burn wound healing potential. Burns 2008;34:761–9.
[23] Jaskille A, Shupp J, Jordan M, Jeng J. Critical review of burn depth assessment techniques: Part I. Historical review. J Burn Care Res 200930:.
[24] Shin J, Yi H. Diagnostic accuracy of laser Doppler imaging in burn depth assessment: systematic review and meta-analysis. Burns 2016;42:1369–76.
[25] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006–12.
[26] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539–49.
[27] Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:22.
[28] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651–7.
[29] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–45.
[30] Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" — a practical guideline. Trials 2016;17:449.
[31] Mihara K, Shindo H, Mihara H, Ohtani M, Nagasaki K, Katoh N. Early depth assessment of local burns by videomicroscopy: a novel proposed classification. Burns 2012;38:371–7.
[32] Mihara K, Nomiyama T, Masuda K, Shindo H, Yasumi M, Sawada T, et al. Dermoscopic insight into skin microcirculation—burn depth assessment. Burns 2015;41:1708–16.
[33] Singer AJ, Relan P, Beto L, Jones-Koliski L, Sandoval S, Clark RA. Infrared thermal imaging has the potential to reduce unnecessary surgery and delays to necessary surgery in burn patients. J Burn Care Res 2016;37:350–5.
[34] Renkielska A, Nowakowski A, Kaczmarek M, Ruminski J. Burn depths evaluation based on active dynamic IR thermal imaging—a preliminary study. Burns 2006;32:867–75.
[35] Renkielska A, Kaczmarek M, Nowakowski A, Grudzinski J, Czapiewski P, Krajewski A, et al. Active dynamic infrared thermal imaging in burn depth evaluation. J Burn Care Res 2014;35:e294–303.
[36] Ruminski J, Kaczmarek M, Renkielska A, Nowakowski A. Thermal parametric imaging in the evaluation of skin burn depth. IEEE Trans Biomed Eng 2007;54:303–12.
[37] Vo LT, Anikijenko P, McLaren WJ, Delaney PM, Barkla DH, King RG. Autofluorescence of skin burns detected by fiber-optic confocal imaging: evidence that cool water treatment limits progressive thermal damage in anesthetized hairless mice. J Trauma 2001;51:98–104.
[38] Mazhar A, Saggese S, Pollins AC, Cardwell NL, Nanney L, Cuccia DJ. Noncontact imaging of burn depth and extent in a porcine model using spatial frequency domain imaging. J Biomed Opt 2014;19:086019.
[39] Maher JR, Jaedicke V, Medina M, Levinson H, Selim MA, Brown WJ, et al. In vivo analysis of burns in a mouse model using spectroscopic optical coherence tomography. Opt Lett 2014;39:5594–7.
[40] Ganapathy P, Tamminedi T, Qin Y, Nanney L, Cardwell N, Pollins A, et al. Dual-imaging system for burn depth diagnosis. Burns 2014;40:67–81.
[41] Zhao Y, Maher JR, Kim J, Selim MA, Levinson H, Wax A. Evaluation of burn severity in vivo in a mouse model using spectroscopic optical coherence tomography. Biomed Opt Express 2015;6:3339–45.
[42] Ida T, Iwazaki H, Kawaguchi Y, Kawauchi S, Ohkura T, Iwaya K, et al. Burn depth assessments by photoacoustic imaging and laser Doppler imaging. Wound Repair Regen 2016;24:349–55.
[43] Mihara K, Shindo H, Ohtani M, Nagasaki K, Nakashima R, Katoh N, et al. Early depth assessment of local burns by videomicroscopy: 24 h after injury is a critical time point. Burns 2011;37:986–93.
[44] McGill DJ, Sorensen K, MacKay IR, Taggart I, Watson SB. Assessment of burn depth: a prospective, blinded comparison of laser Doppler imaging and videomicroscopy. Burns 2007;33:833–42.
[45] Yeong EK, Hsiao TC, Chiang HK, Lin CW. Prediction of burn healing time using artificial neural networks and reflectance spectrometer. Burns 2005;31:415–20.
[46] Iraniha S, Cinat ME, VanderKam VM, Boyko A, Lee D, Jones J, et al. Determination of burn depth with noncontact ultrasonography. J Burn Care Rehabil 2000;21:333–8.
[47] Afromowitz MA, Callis JB, Heimbach DM, DeSoto LA, Norton MK. Multispectral imaging of burn wounds: a new clinical instrument for evaluating burn depth. IEEE Trans Biomed Eng 1988;35:842–50.
[48] Pape SA, Baker RD, Wilson D, Hoeksema H, Jeng JC, Spence RJ, et al. Burn wound healing time assessed by laser Doppler imaging (LDI). Part 1: derivation of a dedicated colour code for image interpretation. Burns 2012;38:187–94.
[49] Monstrey SM, Hoeksema H, Baker RD, Jeng J, Spence RS, Wilson D, et al. Burn wound healing time assessed by laser Doppler imaging. Part 2: validation of a dedicated colour code for image interpretation. Burns 2011;37:249–56.
[50] Nguyen K, Ward D, Lam L, Holland AJ. Laser Doppler Imaging prediction of burn wound outcome in children: is it possible before 48 h? Burns 2010;36:793–8.
[51] La Hei ER, Holland AJ, Martin HC. Laser Doppler imaging of paediatric burns: burn wound outcome can be predicted independent of clinical examination. Burns 2006;32:550–3.
[52] Holland AJ, Martin HC, Cass DT. Laser Doppler imaging prediction of burn wound outcome in children. Burns 2002;28:11–7.
[53] Hoeksema H, Baker RD, Holland AJ, Perry T, Jeffery SL, Verbelen J, et al. A new, fast LDI for assessment of burns: a multi-centre clinical evaluation. Burns 2014;40:1274–82.
[54] Holland AJ, Ward D, La Hei ER, Harvey JG. Laser doppler line scan burn imager (LDLS-BI): sideways move or a step ahead? Burns 2014;40:113–9.
[55] Cho JK, Moon DJ, Kim SG, Lee HG, Chung SP, Yoon CJ. Relationship between healing time and mean perfusion units of laser Doppler imaging (LDI) in pediatric burns. Burns 2009;35:818–23.
[56] Mileski WJ, Atiles L, Purdue G, Kagan R, Saffle JR, Herndon DN, et al. Serial measurements increase the accuracy of laser Doppler assessment of burn wounds. J Burn Care Rehabil 2003;24:187–91.
[57] Stekelenburg CM, Jaspers ME, Niessen FB, Knol DL, van der Wal MB, de Vet HC, et al. In a clinimetric analysis, 3D stereophotogrammetry was found to be reliable and valid for

measuring scar volume in clinical research. J Clin Epidemiol 2015;68:782–7.
[58] De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine. A practical guide. first ed. Cambridge: Cambridge Univ Press; 2011.
[59] de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033–9.
[60] Yeong EK, Mann R, Goldberg M, Engrav L, Heimbach D. Improved accuracy of burn wound assessment using laser Doppler. J Trauma 1996;40:956-61 discussion 61-2.
[61] Bloemen MC, van Zuijlen PP, Middelkoop E. Reliability of subjective wound assessment. Burns 2011;37:566–71.
[62] Jeng JC, Bridgeman A, Shivnan L, Thornton PM, Alam H, Clarke TJ, et al. Laser Doppler imaging determines need for excision and grafting in advance of clinical judgment: a prospective blinded trial. Burns 2003;29:665–70.
[63] Atiles L, Mileski W, Spann K, Purdue G, Hunt J, Baxter C. Early assessment of pediatric burn wounds by laser Doppler flowmetry. Journal Burn Care Rehabil 1995;16:596–601.
[64] Atiles L, Mileski W, Purdue G, Hunt J, Baxter C. Laser Doppler flowmetry in burn wounds. J Burn Care Rehabil 1995;16:388–93.
[65] Riordan CL, McDonough M, Davidson JM, Corley R, Perlov C, Barton R, et al. Noncontact laser Doppler imaging in burn depth analysis of the extremities. J Burn Care Rehabil 2003;24:177–86.