Burn scar assessment: identification and testing of an objective multi-component device


Link to publication in the UWA Research Repository

Rights statement
This work is protected by Copyright. You may print or download ONE copy of this document for the purpose of your own non-commercial research or study. Any other use requires permission from the copyright owner. The Copyright Act requires you to attribute any copyright works you quote or paraphrase.

General rights
Copyright owners retain the copyright for their material stored in the UWA Research Repository. The University grants no end-user rights beyond those which are provided by the Australian Copyright Act 1968. Users may make use of the material in the Repository providing due attribution is given and the use is in accordance with the Copyright Act 1968.

Take down policy
If you believe this document infringes copyright, raise a complaint by contacting repository-lib@uwa.edu.au. The document will be immediately withdrawn from public access while the complaint is being investigated.
Burn scar assessment: Identification and testing of an objective multi-component device

Thilanee Uliya Gankande MD, MHSM
School of Surgery

This thesis is presented for the degree of Doctor of Philosophy at the University of Western Australia

2015
Declaration for the theses containing published work and/or work prepared for publication

The examination of the thesis is an examination of the work of the student. The work must have been substantially conducted by the student during enrolment in the degree.

Where the thesis includes work to which others have contributed, the thesis must include a statement that makes the student’s contribution clear to the examiners. This may be in the form of a description of the precise contribution of the student to the work presented for examination and/or a statement of the percentage of the work that was done by the student.

In addition, in the case of co-authored publications included in the thesis, each author must give their signed permission for the work to be included. If signatures from all the authors cannot be obtained, the statement detailing the student’s contribution to the work must be signed by the coordinating supervisor.

Please sign one of the statements below.

1. This thesis **does not contain** work that I have published, nor work under review for publication.

   Student Signature............................................................................................................................................................

2. This thesis contains **only sole-authored** work, some of which has been published and/or prepared for publication under sole authorship. The bibliographical details of the work and where it appears in the thesis are outlined below.

   Student Signature............................................................................................................................................................

3. This thesis contains published work and/or work prepared for publication, **some of which has been co-authored**. The bibliographical details of the work and where it appears in the thesis are outlined below.

   The student must attach to this declaration a statement for each publication that clarifies the contribution of the student to the work. This may be in the form of a description of the precise contributions of the student to the published work and/or a statement of percent contribution by the student. This statement must be signed by all authors. **If signatures from all the authors cannot be obtained, the statement detailing the student’s contribution to the published work must be signed by the coordinating supervisor.**
Publications

The following published work and work prepared for publication has arisen from the PhD research and comprises Chapters 2, 3 and 4. The reprints of the published works are included as Appendices A, B and C.


Student contribution

For each of the three clinical studies undertaken in Phase Two of the project (Chapter 2, 3 and 4), the student completed the ethics applications, formulated the research protocol and designed the study documentation (patient information sheets, data collection sheets, consent forms etc.) under the guidance of Dr Wallace and Dr Duke. The student participated in each clinical trial as a scar rater.

The student undertook the data entry and data analysis and interpretation for each of the three clinical studies (Chapter 2, 3 and 4). The student prepared the three manuscripts for publication and was involved in the editing of each of the three publications. The contribution to each publication by the student is over 90%.

Contribution of co-authors

Dr Dale Edgar assisted with the ethics application for the clinical study to test the modified Vancouver Scar Scale (mVSS) used at Royal Perth Hospital (Chapter 2). The clinical study was conducted as a sub-study within a quality control project of the scar assessment method (mVSS) for which Dr Edgar was chief investigator.

Prof. Fiona Wood developed the modification of the Vancouver Scar Scale (current standard clinical method of scar assessment at Royal Perth Hospital – mVSS), provided clinical guidance for the research, trained the scar raters used in the first study (Chapter 2) and participated in the third study (Chapter 4) as the expert scar rater.

Dr Janine Duke assisted with experimental design of the clinical studies and with the analysis and interpretation of the data of the three studies (Chapter 2, 3 and 4). Dr Duke also helped with the editing of all the three manuscripts.

Dr Hilary Wallace assisted with experimental design of the clinical studies and supervised the preparation of the study documentation (Chapter 2, 3 and 4). Dr Wallace was involved in the interpretation of data and editing of the three manuscripts.
Anne Henderson (Chapter 2 and 4), Helen DeJong (Chapter 2 and 3), Patricia Danielsen (Chapter 3) and Sarah McGarry (Chapter 4) participated in clinical trials as scar raters.

Presentations
The following presentations have also arisen from this PhD project:


5. Gankande TU, Duke JM, Danielsen PL, DeJong HM, Wood, FM and Wallace HJ. Comparison of objective scar assessments obtained with the DermaLab Combo® with scar assessments of the Modified Vancouver Scar Scale (mVSS). 2013. Australia and New Zealand Burn Association 37th Annual Scientific meeting, Fremantle, Australia (Oral presentation).


Student Signature..............................................................................................................................

Coordinating Supervisor Signature..................................................................................................
Abstract

Scarring is a consequence of burn injury healing and poses a substantial clinical problem. Poor scar outcomes are associated with a wide range of physical, psychological, and aesthetic problems for burn survivors. Currently, there is a lack of objective evidence-based scar assessment methods for clinical scar evaluation. Current scar assessment methods used clinically are subjective and most are low in reliability; objective scar assessment methods are also needed. Therefore this research project was undertaken with an overarching objective to identify and test an objective tool for scar assessment in clinical burns practice.

The work presented in this thesis was undertaken in three phases:

1. An extensive literature review of currently available scar assessment methods was undertaken and identified the DermaLab Combo® (Cortex Technologies, Denmark) multi-component device as an objective measure for testing in clinical burn scar assessment.

2. Three clinical studies (Studies I, II, III) were undertaken to support the testing of the DermaLab Combo® in burn scar assessment.

3. Evaluation of the results, discussion of scar assessment measurement issues and conclusions and recommendations for evidence-based scar assessment and future research.

Study I assessed the inter-rater reliability of the current scar assessment method used at Royal Perth Hospital (RPH), the modified Vancouver Scar Scale (mVSS), to enable its use as reference standard for subsequent testing of the DermaLab Combo®. Thirty subjects with burns scars were recruited and assessed by three raters using the mVSS. The ‘best’ and the ‘worst’ areas of the scar were assessed in comparison to the adjacent normal skin. The weighted kappa statistic (kw) and intra-class correlation coefficient (ICC) were used to measure inter-rater reliability. Inter-rater reliability of mVSS scores were found to be dependent on the severity of the scar, with higher reliability demonstrated for ‘worse’ scar areas. Inter-rater reliability also varied according to the individual scar parameters being measured; pliability demonstrated ‘excellent’ reliability, pigmentation ‘marginal to good’ reliability and both vascularity and height demonstrated ‘good to excellent’ reliability. The potential to misclassify some scar outcomes when using the mVSS was identified.

Study II assessed the reliability (inter-rater and test retest) of the DermaLab Combo® measurements (pigmentation [melanin], vascularity [erythema], pliability and thickness) using an index scar and a matched contralateral normal skin site on 30 subjects. Measurements were made by three raters on three locations within the assessment sites on two occasions. The DermaLab Combo® pigmentation [melanin] measurement demonstrated ‘excellent’ inter-rater and test retest reliability. Vascularity [erythema] measurements had ‘good to excellent’ inter-
rater reliability; however, test-retest reliability was low. The pliability measurements
demonstrated ‘excellent’ test-retest; inter-rater reliability assessment was not possible due to
required wash-out periods. Technical limitations were encountered making measurements of
thickness and pliability using the DermaLab Combo® in some scars. The inter-rater reliability
of pigmentation, vascularity and thickness measurements obtained by the DermaLab Combo®
was greater than that achieved for the mVSS.

Study III explored the interpretation of the DermaLab Combo® scar measurements of
pigmentation [melanin] and vascularity [erythema] with respect to categories similar to the
mVSS for the purpose of data interpretation and harmonisation (Chapter 4). One hundred
subjects were recruited using targeted stratification based on mVSS classification of burn scar
parameters. An index scar and an adjacent normal skin site were assessed using the mVSS and
the DermaLab Combo®. Exploratory analyses were conducted and the concordance of
DermaLab Combo® pigmentation and vascularity measurements in relation to the mVSS was
explored. Classification of pigmentation [melanin] and vascularity [erythema] measurements in
relation to the mVSS was explored. Optimum classification of the DermaLab Combo® colour
measurements into mVSS pigmentation categories was achieved using two predictors – melanin
index % (MI%) and erythema index % (EI%) derived from the DermaLab Combo® melanin
index and erythema index. Vascularity [erythema] classification needs further investigation.

The DermaLab Combo® is an objective multi-component device with high reliability of burn
scar measurements. Assessment of validity at present is hampered by the lack of suitable ‘gold
standard’ for scar assessment. Further testing of the DermaLab Combo® in scar assessment is
needed before the DermaLab Combo® can be translated successfully into clinical burn scar
assessment. The evidence generated from this project provides an initial platform and base to
build future research to enable the translation of the DermaLab Combo® to clinical burn scar
assessment in the future.
Table of Contents

Declaration for the theses containing published work and/or work prepared for publication .... iii
Abstract....................................................................................................................................... vii
Table of Contents.......................................................................................................................... ix
List of Tables .................................................................................................................................... xv
List of Figures ..................................................................................................................................... xvii
Acknowledgements ..................................................................................................................... xix
List of Abbreviations .................................................................................................................. xxi
Overview and aims of the project ............................................................................................ xxiii
Chapter 1: Literature review ...................................................................................................... 1
  1.1 Epidemiology of post-burn scarring ........................................................................ 3
    1.1.1 Prevalence .......................................................................................................... 3
    1.1.2 Impact of improved patient survival on scarring ............................................... 7
    1.1.3 Costs related to scarring ..................................................................................... 7
  1.2 Post-burn scarring – definitions and classification .................................................... 8
    1.2.1 Scar phenotype ................................................................................................... 8
    1.2.2 Classification of post-burn scarring ................................................................... 8
  1.3 Mechanisms of scarring ............................................................................................. 9
    1.3.1 Overview of wound healing stages .................................................................... 9
    1.3.2 Burn injury wounds .......................................................................................... 10
    1.3.3 Scarring and wound repair ............................................................................... 10
      1.3.3.1 Pigmentation ................................................................................................ 10
      1.3.3.2 Vascularity ................................................................................................... 11
      1.3.3.3 Fibrosis ......................................................................................................... 12
  1.4 Impact of scarring after burn injury ......................................................................... 12
    1.4.1 Cosmetic impacts ............................................................................................. 12
    1.4.2 Functional impacts ........................................................................................... 13
    1.4.3 Psychosocial impacts ....................................................................................... 14
  1.5 Assessment of post-burn scarring ............................................................................ 14
    1.5.1 Context of scar assessment .............................................................................. 14
3.2.1 Subjects ............................................................................................................ 58
3.2.2 Study design ..................................................................................................... 58
3.2.3 Raters ............................................................................................................... 59
3.2.4 Materials .......................................................................................................... 59
  3.2.4.1 The colour probe .......................................................................................... 59
  3.2.4.2 The elasticity probe ...................................................................................... 59
  3.2.4.3 The skin thickness probe .............................................................................. 60
3.3 Procedure ................................................................................................................. 60
  3.3.1 Inter-rater reliability ......................................................................................... 60
  3.3.2 Test-retest reliability ........................................................................................ 62
3.4 Data collection and analysis ..................................................................................... 62
  3.4.1 Inter-rater reliability ......................................................................................... 63
  3.4.2 Test-retest reliability ........................................................................................ 63
3.5 Results ................................................................................................................... 64
  3.5.1 Subjects and descriptive statistics .................................................................... 64
  3.5.2 Inter-rater reliability ......................................................................................... 65
  3.5.3 Test-retest reliability ........................................................................................ 67
3.6 Discussion ................................................................................................................ 68
3.7 Conclusion ............................................................................................................... 71

Chapter 4: Exploring the interpretation of the DermaLab Combo® pigmentation and
vascularity measurements via the mVSS ................................................................................. 73
4.1 Introduction ............................................................................................................ 75
4.2 Method .................................................................................................................... 76
  4.2.1 Subjects ............................................................................................................ 76
  4.2.2 Study design ..................................................................................................... 77
  4.2.3 Raters ............................................................................................................... 77
4.3 Scar assessment ...................................................................................................... 77
  4.3.1 The mVSS (Baryza and Baryza modification)....................................................... 77
  4.3.2 The DermaLab Combo® device – colour probe .............................................. 77
4.4 Data collection and procedure .................................................................................. 78
4.5  Analysis and sample size ......................................................................................... 79
  4.5.1  Generation of melanin index % (MI%) and erythema index % (EI%) ............ 80
  4.5.2  Exploratory data analysis ................................................................................. 80
4.6  Results ................................................................................................................... .. 81
  4.6.1  Subject stratification and descriptive statistics ................................................ 81
  4.6.2  The mVSS rater agreement .............................................................................. 82
  4.6.3  Exploratory data analysis – pigmentation ........................................................ 82
  4.6.4  Exploratory data analysis – vascularity ........................................................... 86
4.7  Discussion................................................................................................................ 89
4.8  Conclusion ............................................................................................................... 91

Chapter 5: Evaluation of the clinical studies, general discussion, conclusion and
recommendations ............................................................................................................. 93

  5.1  Evaluations of the clinical studies ................................................................. 95
    5.1.1  Reliability of the mVSS (Study I – Chapter 2) ............................................ 95
      5.1.1.1  Summary of study .............................................................................. 95
      5.1.1.2  Measurement issues ......................................................................... 95
      5.1.1.3  Conclusions ....................................................................................... 97
    5.1.2  Reliability of the DermaLab Combo® (Study II – Chapter 3) .................... 97
      5.1.2.1  Summary of study .............................................................................. 97
      5.1.2.2  Measurement issues ......................................................................... 97
      5.1.2.3  Conclusion ........................................................................................ 99
    5.1.3  Exploration of the DermaLab Combo® pigmentation and vascularity
          measurements in relation to the mVSS (Study III – Chapter 4) ..................... 100
      5.1.3.1  Summary of study .............................................................................. 100
      5.1.3.2  Methodological issues ...................................................................... 100
      5.1.3.3  Conclusions ....................................................................................... 101
  5.2  Future testing of the DermaLab Combo® – considerations ............................ 101
    5.2.1  Standardised protocols .............................................................................. 101
    5.2.2  Improved experimental design .................................................................... 102
      5.2.2.1  Homogeneity of study subjects ......................................................... 102
      5.2.2.2  Sample size ....................................................................................... 102
List of Tables

Chapter 1
Table 1-1: Prevalence information on post-burn hypertrophic scarring ........................................ 5
Table 1-2: Clinically observable changes in hypertrophic scars during maturation .................... 12
Table 1-3: Published studies of scar assessment scales ............................................................... 18
Table 1-4: Summary of devices that measure scar colour categorised by the principle of measurement ................................................................................................................................ 29
Table 1-5: Summary of devices that measure scar thickness/height categorised by the principle of measurement ............................................................................................................................ 31
Table 1-6: Summary of devices that measure scar pliability/elasticity ........................................ 32

Chapter 2
Table 2-1: Numerical scale of mVSS used in the study............................................................... 44
Table 2-2: Weighted Kappa (k_w) values for individual rater pairs in 'best' and 'worst' scar areas 48
Table 2-3: ICC and Bland Altman values (mean score difference, 95% limits of agreement) for total mVSS for “best” and ‘worst’ area of scar for rater pairs ................................................................. 49
Table 2-4: ICC and Bland Altman values (mean score difference, 95% limits of agreement) for scar %TBSA and allocation of scar %TBSA to <5 mVSS and 5-10 mVSS categories for rater pairs ............................................................................................................................ 50
Table 2-5: Clinically significant misclassification rates of individual rater pairs for ‘best’ and ‘worst’ areas of the scar .................................................................................................................. 50

Chapter 3
Table 3-1: Range of scar measurements at baseline for pigmentation, vascularity, thickness and pliability measured with the DermaLab Combo® for the 30 study subjects ............................... 64
Table 3-2: Inter-rater reliability for pigmentation, vascularity and thickness components based on individual site measurements (all rater combinations) ................................................................. 66
Table 3-3: Inter-rater reliability for pigmentation, vascularity and thickness components using average measurements of each area (all rater combinations) .......................................................................................................... 68
Table 3-4: Test-retest reliability for pigmentation, vascularity and thickness components using individual site measurements ................................................................................................................. 68

Chapter 4
Table 4-1: The number of subjects recruited using the mVSS (R1) and absolute agreement between R1 and R3 used in the analyses ........................................................................................................... 82
Table 4-2: Descriptive statistics for MI% values for mVSS pigmentation categories ......................... 83
Table 4-3: Classification concordance results for MI% by mVSS pigmentation categories for cut-off values using standard and transformed X axis................................................................. 86
Table 4-4: Descriptive statistics for EI% values for mVSS vascularity categories............... 87

Chapter 5
Table 5-1: Key results of mVSS reliability testing................................................................. 95
Table 5-2: Key results of reliability testing of the DermaLab Combo® ................................. 97
Table 5-3: Interpretation of DermaLab Combo® pigmentation and vascularity measurements by mVSS classification................................................................................................. 100
Table 5-4: Project summary – inter-rater reliability the DermaLab Combo® and the mVSS .. 105
List of Figures

Chapter 2
Figure 2-1: Comparison of scores using conventional and new method mVSS-TBSA .......... 42
Figure 2-2: Allocated mVSS scores by raters in ‘best’ and ‘worst’ areas of the index scar ....... 47

Chapter 3
Figure 3-1: The DermaLab Combo® device and example outputs for the four parameters – melanin, erythema, thickness and elasticity (adapted from www.cortextechnologies.dk) .......... 60
Figure 3-2: Sampling protocol within marked 3 cm x 3 cm square used for measurement...... 61

Chapter 4
Figure 4-1: The DermaLab Combo® colour probe and display ................................................. 78
Figure 4-2: Probe placement for the DermaLab Combo® measurement within the 3 cm x 3 cm scar and the normal skin areas ................................................................. 79
Figure 4-3: Dot plot of MI% scores of scars for pigmentation by mVSS with summary statistics (median, minimum-maximum) .................................................................................................................. 84
Figure 4-4: MI% and EI% values by mVSS pigmentation category (A) Cut-off values on the X axis (MI%) (B) Cut-off values on the transformed axes ........................................................................................................ 85
Figure 4-5: Dot plot of EI% values of scars for vascularity by mVSS with summary statistics (median, minimum-maximum) ............................................................................................................................ 88
Figure 4-6: EI% and MI% values by mVSS vascularity category .............................................. 88
Acknowledgements

I would firstly like to express my appreciation and thanks to my funding sources: The Wound Management and Innovation CRC, Australian Post-graduate Award (APA) and the University of Western Australia for my scholarships enabling me to take on this PhD project.

I wish to say a special thank you to my supervisors Hilary, Janine and Fiona. To Hilary, thank you for being available to me every step of the way throughout my PhD project. Your attention to detail and feedback was very helpful. You not only encouraged my PhD research but also encouraged my involvement in other areas of research within the field of burns allowing me to grow as a researcher. To Janine, I say a big thank you for your guidance, encouragement and constructive feedback. You were available to me whenever I needed guidance, encouragement and support to get this thesis completed on time and for that I am forever grateful. Your knowledge in data analysis and interpretation expanded my interest in data exploration and dare I say I have learned to ‘enjoy’ the whole process of data analysis! I also thank you for the kindness you have shown me during my daughter’s illness which took a lot of my energy but your gentle encouragement saw me through many difficult days. To Fiona, thank you for the clinical guidance you have given to my project and participating in my clinical trial despite a very busy schedule, your optimism and passion was infectious and was helpful when patient recruitment seemed untenable!

I also want to say thank you Dale my external supervisor for supporting me though the first clinical study ethics process. I thank Delia Hendrie for her support during the early stages of my research project and a big thank you goes out to Jeremy Wallace for coming to my rescue in instances when I faced statistical glitches!

I thank Anne Henderson, Helen DeJong, Patricia Danielsen and Sarah McGarry for participating in my three clinical trials alongside me as scar raters. Without your support I could not have completed the studies. I say a big thank you to Suzanna Rea and the other doctors at the Royal Perth Hospital Burns Clinic and to Joy, Sharon, Maggie and the nursing staff for their support when I recruited patients and collected data for my PhD project.

Thank you to all of my friends at the Burns Injury Research Unit, Mark, Andy, Mansour, Vetri, Dulharie and Mitali for supporting me in numerous ways. Thank you Dulharie for your help with the graphs and Mansour your coffee is what kept me going on many difficult mornings!

I thank my friend Brenda for her help with the formatting of my thesis and for being there whenever I need help every single day! Brenda I say thank you from the bottom of my heart.

Last but not least a heartfelt thank you to my family. Words cannot express how grateful I am to my husband Sugeeesh, the sacrifices that you’ve made on my behalf have allowed me to pursue my dream! Thank you and I love you more than you will ever know. I say thank you to my daughters Aneeka and Shani for putting up with a busy mummy for three long years, I love you both very much and now mummy will have more time to nag you now.

xix
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the curve</td>
</tr>
<tr>
<td>AU$</td>
<td>Australian dollar</td>
</tr>
<tr>
<td>BA</td>
<td>Bland-Altman method</td>
</tr>
<tr>
<td>CIE</td>
<td>Commission Internationale de l'Eclairage</td>
</tr>
<tr>
<td>dB</td>
<td>Decibels</td>
</tr>
<tr>
<td>EI%</td>
<td>Percentage change in erythema relative to normal skin</td>
</tr>
<tr>
<td>GB</td>
<td>Gigabyte</td>
</tr>
<tr>
<td>Hb</td>
<td>Haemoglobin</td>
</tr>
<tr>
<td>HS</td>
<td>Hypertrophic scar</td>
</tr>
<tr>
<td>HSS</td>
<td>Hamilton Scar Scale</td>
</tr>
<tr>
<td>ICC</td>
<td>Intra-class Correlation Coefficient</td>
</tr>
<tr>
<td>IQR</td>
<td>Inter-quartile range</td>
</tr>
<tr>
<td>$k_w$</td>
<td>Weighted kappa statistic</td>
</tr>
<tr>
<td>L</td>
<td>Luminance</td>
</tr>
<tr>
<td>LED</td>
<td>Light emitting diode</td>
</tr>
<tr>
<td>MAPS</td>
<td>Matching Assessment of Scars and Photographs</td>
</tr>
<tr>
<td>MHz</td>
<td>Megahertz</td>
</tr>
<tr>
<td>MI</td>
<td>Melanin index</td>
</tr>
<tr>
<td>MI%</td>
<td>Percentage change in melanin relative to normal skin</td>
</tr>
<tr>
<td>mm</td>
<td>Millimetres</td>
</tr>
<tr>
<td>MPa</td>
<td>Megapascals</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>mRSS</td>
<td>Modified Rodnan Skin Sore</td>
</tr>
<tr>
<td>MSS</td>
<td>Manchester Scar Scale</td>
</tr>
<tr>
<td>mVSS</td>
<td>Modified Vancouver Scar Scale</td>
</tr>
<tr>
<td>mVSS-TBSA</td>
<td>Modified VSS-TBSA method</td>
</tr>
<tr>
<td>N, n</td>
<td>Number</td>
</tr>
<tr>
<td>nd</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>nm</td>
<td>Nanometer</td>
</tr>
<tr>
<td>OT</td>
<td>Occupational therapy</td>
</tr>
<tr>
<td>PASI</td>
<td>Psoriasis Area and Severity Index</td>
</tr>
<tr>
<td>POSAS</td>
<td>Patient and Observer Scar Assessment Scale</td>
</tr>
<tr>
<td>Pn</td>
<td>Pigmentation measurement for normal skin</td>
</tr>
</tbody>
</table>
Ps Pigmentation measurement for scar
QC Quality control
R1, R2, R3 Scar raters 1, 2 and 3
RAM Random access memory
ROC Receiver Operating Characteristics
RPH Royal Perth Hospital
SBSES Stony Brook Scar Assessment Scale
SPSS Statistical Package for the Social Sciences
SSS Seattle Scar Scale
TBSA Total body surface area
TEWL Trans epidermal water loss
TUPS Tissue ultrasound palpation
UK United Kingdom
USA United States of America
USB Universal serial bus
US$ American dollar
UVR Ultraviolet radiation
UWA University of Western Australia
VAS Visual Analogue Scale
VSS Vancouver Scar Scale
Vn Vascularity measurement for normal skin
Vs Vascularity measurement for scar
WA Western Australia
WHO World Health Organisation
2SD 2 standard deviations
ΔP Percentage change in pigmentation
ΔV Percentage change in vascularity
τb Kendall Tau-b statistic
Overview and aims of the project

Burn injuries are a significant health problem worldwide (World Health Organisation (WHO) 2011) and burn injury wounds remain one of the most complex acute wounds to treat (Pham et al. 2007). Post-burn survival rates have improved substantially over recent decades, due to the progress in the field of intensive care management, with improved management of shock and antimicrobial therapies, and better nutritional support (Van der Wal 2013; Cleland 2006; Druery et al. 2005; Mayhall 2003; Lootens et al. 2013; Song & Chua 2005; Gangemi et al. 2008). With increasing survival of burn patients there has been a major shift in focus from mortality to the longer-term health and quality of life outcomes of the burn survivor (Druery et al. 2005; Cleland 2006; Song & Chua 2005; Gangemi et al. 2008).

Scarring is a consequence of burn injury healing and poses a substantial clinical problem (Gangemi et al. 2008). Poor scar outcomes are associated with a wide range of physical, psychological, and aesthetic problems for burn survivors (Wiechman & Patterson 2008; Bloemen et al. 2011a; Gangemi et al. 2008). Many burn patients are at risk of long-term morbidity due to scarring, leading to reduced quality of life (Baur et al. 1998; Rumsey et al. 2003; Van Loey & Von Son 2003; Gangemi et al. 2008). Despite the advances in burn care over recent decades scarring remains a problem for clinicians and burn survivors (Sheridan & Thompkins 2004). The prevalence of post burn scarring is estimated to be between 30% and 70% (Gangemi et al. 2008; Bombaro et al. 2003).

Evidence-based scar management at present is challenging due to limitations of currently available validated tools to evaluate scar outcome and a lack of well-designed studies. Within the published literature there are only a few data-supported studies available on post-burn scar behaviour and on predictors for severe scarring (Van Zuijlen et al. 2002; Nguyen et al. 2008; Brusselaers et al. 2010a).

Scar assessment is essential for diagnosis, monitoring, evaluation and therapeutic management (Bayat et al. 2003). Several scar assessment scales have been proposed and are routinely used within the clinical setting to evaluate post-burn scarring and the response of scars to therapy. The Vancouver Scar Scale (VSS) proposed by Sullivan in 1990 (Sullivan et al. 1990), together with its subsequent modifications, have been used extensively in clinical scar assessment and as an outcome measure in burn studies (Tyack et al. 2012). Scar assessments conducted using scar measurement scales are subjective and highly observer dependent (Draaijers et al. 2004c). At present there are some devices that are used to obtain objective assessment of individual scar parameters. The devices described in published literature assess scar parameters including pliability, firmness, colour, perfusion, thickness, and three-dimensional topography (Fearmonti et al. 2010). These devices provide a quantitative measure of a single physical scar attribute in an objective manner. However, there is still a lack of consensus regarding the most appropriate
and clinically applicable evaluation instrument that provides objective scar assessment for all the important scar parameters (Fearmonti et al. 2010). At present there is no accepted ‘gold standard’ method in scar assessment.

The lack of objective scar assessment methods for scar evaluation suggests the need for a reliable assessment technique. A scar measuring device which is non-invasive, reliable, accurate, reproducible, easy-to-use and provides objective scar measurements for all important scar parameters is required. A device which is affordable and is commercially available would have great clinical utility.

Aims

The overarching aim of the project was to translate an objective scar assessment method into a new tool for clinical burns practice. The specific objectives of the project were:

1. To identify an objective device with the capacity to measure multiple scar components analogous to those measured by the Vancouver Scar Scale (VSS)

2. To assess the inter-rater reliability of a modified Vancouver Scar Scale (mVSS) used at Royal Perth Hospital (RPH)

3. To assess the inter-rater and test-retest reliability of the objective scar assessment device identified in (1)

4. To explore the interpretation and validity of measurements obtained with the objective scar assessment device in categories similar to the mVSS

In order to achieve the specific project objectives, the research was conducted in three phases. The first phase involved a comprehensive literature review of scar assessment methods currently available to identify a suitable device capable of measuring multiple scar components analogous to the VSS (pigmentation, vascularity, pliability and height) [Chapter 1]. The second phase of the project consisted of three separate clinical studies. The first study tested the inter-rater reliability of the mVSS currently used at RPH. The findings of this study [Chapter 2] were published in Burns (Gankande et al. 2013) prior to the submission of this thesis. The second study tested the inter-rater and test-retest reliability of the DermaLab Combo®, the multi-component skin-testing device identified in phase one. The findings of this study [Chapter 3] were also published in Burns (Gankande et al. 2014) prior to submission of this thesis. The third and final clinical study investigated the interpretation of the pigmentation (melanin) and vascularity (erythema) measurements of burns scars obtained with the DermaLab Combo® device in categories analogous to mVSS and to generate evidence to establish validity. The findings of this study [Chapter 4] too was accepted for publication in Burns (Gankande et al. 2015) prior to the submission of this thesis. The final phase of the project summarises the results.
of the clinical studies and links the research findings to discuss potential evidence-based
directions in scar assessment. Conclusions are drawn and recommendations for future clinical
scar assessment trials using the DermaLab Combo® device are made [Chapter 5].

Approval for this research project was granted by the Human Research Ethics Committees of
the University of Western Australia (UWA) and RPH.
Chapter 1
Chapter 1 – Literature Review

This chapter describes the epidemiology and mechanisms of post-burn scarring, the clinical features of burn scars, subjective and objective burn scar assessment tools and the rationale for testing an objective multi-component device for potential use in clinical burn scar assessment.

1.1 Epidemiology of post-burn scarring

1.1.1 Prevalence

Prevalence estimates of post-burn scarring vary from 40% to 91% in the published literature (Bombaro et al. 2003; Bloemen et al. 2009). Table 1-1 summarises eight studies that report the prevalence of post-burn scarring. The wide range in the prevalence estimates reported may be attributable to a range of factors including the definition of scar outcomes applied in each study, methods of assessment, age and ethnicity of the study populations used, time since injury and study sample sizes.

All studies presented in Table 1-1 were retrospective patient record audits involving a single centre and used different definitions of scar. Three studies used the inclusion of the word ‘hypertrophic scar’ in the patient record as evidence of scarring (Spurr & Shakespeare 1990; Bombaro et al. 2003; Cubison et al. 2006) and four studies did not specify the definition of scar outcome used (Deitch et al. 1983; McDonald & Deitch 1987; Lewis & Sun 1990; Dedovic et al. 1999). The remaining study (Gangemi et al. 2008) used the definition of scar described by Magliacani and others (1997) of ‘hypertrophy, hypertrophy and contracture, contracture and atrophy’.

In half of the studies the method of scar assessment used was not described (Bombaro et al. 2003; Dedovic et al. 1999; Spurr & Shakespeare 1990; Deitch et al. 1983) and it is possible that scar assessment was based on clinical opinion. Cubison and others (2006) and Gangemi and others (2008) relied on clinical opinion for assessing scars. Only two studies (McDonald & Deitch 1987; Lewis & Sun 1990) described the use of a scar assessment scale; however, these scales are not similar to any of the scar assessment scales in published literature.

The study method and sample size may have contributed to the variations of reported prevalence. Seven of the eight studies were retrospective and it is common to encounter many sources of error due to measurement and confounding factors in retrospective studies. Most studies (refer to Table 1-1) involved small sample sizes of between 60 to 100 subjects (Deitch et al. 1983; McDonald & Deitch 1987; Spurr & Shakespeare 1990; Dedovic et al. 1999; Bombaro
et al. 2003). Studies with small sample sizes are less reliable and have low statistical power to
detect the true prevalence. Only three studies (Cubison et al. 2006; Dedovic et al. 1999;
Gangemi et al. 2008) reported estimates using sample sizes of over 300 patients.

The studies presented in Table 1-1 represent a spectrum of research with heterogeneous
hypotheses regarding the risk of scarring, with consequent recruitment of different study
populations to explore or explain variations. Some studies examined skin colour, for example,
scarring among black vs. white burn patients in United States America (USA) (Deitch et al.
1983; McDonald & Deitch 1987; Bombaro et al. 2003) concluded that white populations were
less prone to scarring (white 7% to 75% compared to black/non-white 31% to 75%). Lewis and
Sun (1990) investigated the prevalence of scarring in a Chinese population (n=58 subjects)
found a very high prevalence rate for hypertrophic scarring (91.4%). The sample sizes of these
studies were approximately 70 to 100 subjects and ability to detect the true prevalence of
scarring was low. Some authors focused on age and examined prevalence among adult and
paediatric populations (Deitch et al. 1983; McDonald & Deitch 1987). The prevalence among
children ranged from 13% to 75% while the adults showed 17% to 75% prevalence of scarring.

Four studies (Deitch et al. 1983; McDonald & Deitch 1987; Cubison et al. 2006; Gangemi et al.
2008) explored prevalence of scarring in relation to time to healing. All four studies
demonstrated an increased risk of scarring when healing took more than 14 days (prevalence
estimates between 52% and 78%) compared to shorter healing times of less than 14 days
(prevalence estimates 2% to 37%). Some authors (Spurr & Shakespeare 1990; Cubison et al.
2006) compared the prevalence of scarring in burn wounds that healed spontaneously to burn
wounds that healed post-burn surgery. Both authors reported high prevalence estimates of post-
surgical scarring (59% to 65%). Gangemi and others (2008) identified that in burns patients the
risk of hypertrophic scarring doubled for every surgical operation performed irrelevant of the
type of surgical intervention.
<table>
<thead>
<tr>
<th>Authors &amp; Country of study</th>
<th>Study design</th>
<th>Population [ethnicity]</th>
<th>Sample size</th>
<th>Definition of scar/ method of scar assessment</th>
<th>Prevalence estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deitch et al. (1983) USA</td>
<td>Retrospective audit of medical records Single centre</td>
<td>Adults; children &lt;14 years [black: white] Adults (n=41) Children (n=59)</td>
<td>Definition and scar assessment method not specified</td>
<td>30% adults (black) 15% adults (white) 31% children (black) 13% children (white) Stratified by time to healing &gt;21 days 78% (for both adult and children – black and white)</td>
<td></td>
</tr>
<tr>
<td>McDonald and Deitch (1987) USA</td>
<td>Retrospective audit of medical records Single centre</td>
<td>Adults; children &lt;14 years [black: white] Adults (n=44) Children (n=26)</td>
<td>Definition not specified Assessment method: Excellent = flat (not thickened) Good ≤ 5% thickened or elevated Fair ≥ 5% thickened or elevated Poor (P) = elevated &gt; 2 mm or contracture</td>
<td>50% adults (black) 17% adults (white) 75% children (black &amp; white) Stratification by time to healing: &lt;14 days = 24% and &gt;21 days = 37% (adult and children – black and white)</td>
<td></td>
</tr>
<tr>
<td>Spurr and Shakespeare (1990) UK</td>
<td>Retrospective audit of medical records Single centre</td>
<td>Children &lt;5 years [not indicated] Two time periods 1968 (n=70) 1984 (n=82)</td>
<td>Definition = the inclusion of the word ‘hypertrophic scar’ in patient record Scar assessment method not specified</td>
<td>65% (1968 and 1984)</td>
<td></td>
</tr>
<tr>
<td>Lewis and Sun (1990) China</td>
<td>Retrospective review Single centre</td>
<td>Adults (age not indicated) [Chinese] Adults (n=58)</td>
<td>Definition not specified Assessment method: Scar grade (0-4) with significant hypertrophy being &gt;1</td>
<td>91.4% significant hypertrophy</td>
<td></td>
</tr>
<tr>
<td>Authors &amp; Country of study</td>
<td>Study design</td>
<td>Population [ethnicity]</td>
<td>Sample size</td>
<td>Definition of scar/ method of scar assessment</td>
<td>Prevalence estimates</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>-------------</td>
<td>-----------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Bombaro et al. (2003)</td>
<td>Retrospective audit of medical records Single centre</td>
<td>Adults (mean age 33 years) [white: non-white]</td>
<td>Adults (n=89)</td>
<td>Definition = the inclusion of the word ‘hypertrophic scar’ in patient record Scar assessment method not specified</td>
<td>75% (non-white) 60% (white)</td>
</tr>
<tr>
<td>Cubison et al. (2006)</td>
<td>Retrospective audit of medical records Single centre</td>
<td>Children &lt;5 years [not specified]</td>
<td>Children (n=337)</td>
<td>Definition = the inclusion of the word ‘hypertrophic scar’ in patient record Assessment – clinical opinion of burns surgeon</td>
<td>35% (overall) Stratified by time to heal 10-14 days 8%; 22-30 days 52% Stratified by treatment Conservative: 10-14 days 2%; 22-30 days 45% Surgery: 10-14 days 33%; 22-30 days 59%</td>
</tr>
<tr>
<td>Gangemi et al. (2008)</td>
<td>Retrospective audit of medical records Single centre</td>
<td>Adults (25-54 years) [not specified]</td>
<td>Adults (n=703)</td>
<td>Scar defined using Magliacani et al. (1997) method Assessment – clinical opinion of burns surgeon</td>
<td>72% adults Risk of scarring doubled for every surgical operation performed</td>
</tr>
</tbody>
</table>
1.1.2 Impact of improved patient survival on scarring

Post-burn survival rates for severe burn injury have improved considerably over recent decades in developed countries (Gangemi et al. 2008; Lootens et al. 2013; Van der Wal et al. 2012b). During the post-World War II era, approximately 50% of patients with burns involving 40% total body surface area (TBSA) survived; however, since the late 1990s, more than 50% of patients with burn injuries involving greater than 80% TBSA now survive in developed countries (Saffle 1998). The survival of patients with large %TBSA burns brings about a high potential for poor scar outcomes and scarring ‘remains a terrible clinical problem’ (Sheridan & Thompkins 2004). Scar quality has now developed into one of the key focus areas in burn care.

1.1.3 Costs related to scarring

The costs of burn injury are incurred by the health care system and the patient. With respect to the health care system costs, approximately US$18 billion is spent annually treating patients admitted to specialised burn centres in the US (Church et al. 2006). It is estimated that the initial cost of hospitalisation per burns patient in the US exceeds US$200,000. However, the overall cost per burns patient is much greater when repeat admissions for rehabilitation and reconstruction are taken into account (Church et al. 2006). These additional costs are directly related to post-burn scarring. In Australia, the significant financial cost of post-burn scarring was reported in the Roundtable Forum on Burns Prevention, House of Representatives Standing Committee on Health and Ageing in 2010 (The Parliament of the Commonwealth of Australia 2010). It was estimated that the AU$65 million per year (2007-08) to treat acute burns in Australia is one-third of the true cost because treatment is not a one-off intervention.

The costs to burn survivors are both financial and social. Many burns survivors face a lifetime of ongoing treatment and rehabilitation due to scarring. Ongoing financial costs are compounded for some families because family members may have to leave their paid employment. A significant social cost of scarring remains the impact on the mental and emotional well-being of the burns survivor and their immediate family. In a briefing to the Committee prior to the Forum (November 2009), Professor Fiona Wood indicated that burn injuries should be considered a chronic disease due to the long-term social and health impacts on the individual, their family and the Australian community (Personal Communication).
1.2 Post-burn scarring – definitions and classification

1.2.1 Scar phenotype

There are a range of definitions of scarring in the published literature depending on the context. Ferguson and others (1996 p.854) define scarring in general as a ‘macroscopic disturbance of the normal structure of the skin architecture, resulting from the end product of a healed wound’. Mustoe and others (2002 p.561) have defined a post-burn scar to be a ‘widespread, red, raised itchy scar that remains within the borders of the burn injury’.

Clinically, scar tissue differs from normal skin by an abnormal colour, increased thickness, irregular skin surface area, and poor functional ability (Lammers 2011). Each of these individual scar features contributes to the final form and cosmetic acceptability of a scar which can be very different from the normal skin. Histologically, scar tissue is associated with increased vascularisation, changes in the normal pattern of pigmentation and sustained increased fibrosis (Aarabi et al. 2007).

1.2.2 Classification of post-burn scarring

A normotrophic scar is considered to be the end-point of the normal process of wound repair (Gangemi et al. 2008). It is well-known that some scars evolve in time very differently from the normal process of wound repair (Maurizio et al. 2008; Satish & Kathju 2010). As a result of abnormal healing the wound is repaired with non-functional tissue which is highly vascular and filled with inflammatory cells and fibroblasts leading to a pathological scar (Singer & Clark 1999). Pathological scars include widespread scars, atrophic scars, scar contractures, hypertrophic scars (HS), and keloid scars (Mustoe et al. 2002).

Hypertrophic scarring is the most common type of pathological scar after burn injury (Bombaro et al. 2003; Anzarut et al. 2005). A commonly used definition for hypertrophic scar is that a hypertrophic scar is raised above the skin level and stays within the confines of the original lesion (Bloemen et al. 2009; Kose & Waseem 2007; Church et al. 2006). Clinically, hypertrophic scars are painful, and some of the main features are: the elevation above skin surface limited to injury borders, redness, itching, prone to contractures and regress within 12 to 24 months. This may lead to the loss of normal skin function and also contribute to disfiguring appearance (Amadeu et al. 2003; Wang et al. 2010).
1.3 Mechanisms of scarring

Wound repair is vital for survival. In mammals, including humans, wound healing occurs via a process of regeneration at early embryonic stages resulting in scar-free wound healing (Heng 2011; Ferguson & O'Kane 2004; Martin 1997). However, wounds in adults and children heal through the process of wound repair rather than regeneration resulting in scarring (Martin 1997; Ferguson & O'Kane 2004; Heng 2011; Ehrlich 1998). Therefore, scar formation is the end point of the natural process of wound repair which is initiated upon injury (Ehrlich 1998; Miller & Nanchahal 2005).

1.3.1 Overview of wound healing stages

Wound healing is an active process involving many organs and systems of the human body and consists of three phases that overlap in time: firstly, the inflammatory phase, secondly, the tissue formation phase and finally, the tissue remodelling phase (Singer & Clark 1999).

Immediately upon injury the inflammatory phase is initiated. This phase consists of two stages namely the vascular response and the cellular response and involves the influx of inflammatory cells and release of mediators to prevent infection and to initiate wound repair. The formation of a blood clot in the wound provides haemostasis and a provisional extracellular matrix for cell migration. Numerous growth factors and mediator pathways are activated to facilitate the secretion of the new dermal matrix (Singer & Clark 1999).

The second phase (tissue formation phase), begins hours after the injury and in parallel to the inflammatory phase and sees the formation of new tissues. Alongside matrix deposition, the process of revascularisation and re-innervation of the tissue begins (Singer & Clark 1999). The new dermal matrix, consists primarily of collagen and is a foundation for the new epithelium to grow (Singer & Clark 1999).

The final phase of the wound healing is the phase of remodelling, which may last for up to 12 months in most wounds; however, in burn wounds it may be prolonged up to 24 months (Maurizio et al. 2008). The abnormal appearance and texture of the scar decreases to some degree during the remodelling phase. Despite this prolonged remodelling, the mature scar may remain very different to normal skin, with fewer or no adnexal structures and may have diminished function (Singer & Clark 1999; Van Zuijlen 2002).
1.3.2 Burn injury wounds

Burn wounds can occur in any anatomical part of the body and may cover large surface areas making them among the largest wounds encountered in surgical practice (Maurizio et al. 2008). Extensive damage to the skin surface (i.e. larger TBSA burns) has the potential to include areas of deep thickness wounds that can cause massive damage to the underlying structures (i.e. deep burn wounds) which in turn may lead to severe scarring (Heng 2011). Dunkin and others (2007) demonstrated the association between scarring and the depth of the burn injury.

During the wound healing process the physiological response of wound contraction is greater in a burn wound compared to other types of wounds due to the larger surface areas involved, and may lead to hypertrophic scarring with contracture in such instances (Junker et al. 2008). Wounds located in areas of increased mobility (i.e. shoulders, pre-sternal chest and clavicular areas), causing constant repeat tension during the wound repair process are prone to developing hypertrophic scarring (Heng 2011).

1.3.3 Scarring and wound repair

While there is no scientific consensus on the patho-physiology of post-burn scarring (Gangemi et al. 2008; Miller & Nanchahal 2005), there appear to be several stages of healing significantly altered in scarring, from the inflammatory phase to the remodelling phase (Satish & Kathju 2010). The effects of these altered stages of healing are seen in clinically measurable components of scars such as altered patterns of pigmentation, increased vascularity and fibrosis which manifests as changes in scar thickness and/or abnormal pliability of scar tissue.

1.3.3.1 Pigmentation

There is a lack of published literature related to the re-pigmentation of scars following cutaneous injury despite extensive research done in recent years into both pigment cells and wound healing (Chadwick et al. 2012). Disruption to normal melanogenesis during the wound repair process will result in abnormal pigmentation in the resulting scar (Tyack et al. 1997). Abnormal pigmentation post-injury is due to changes in the epidermal melanin and the vascular response. Changes in the epidermal melanin content are likely to be an indication of local damage (Tyack et al. 1997). Chadwick and others (2012) hypothesise that re-pigmentation of scars depend on the availability of melanocytes and on the mechanism of injury. The authors further indicate that in partial thickness injuries where wounds contain residual adnexal elements have the capacity to provide melanocytes and epithelial cells to the neo-epithelium and have fewer issues with pigmentation (Chadwick et al. 2012). Clinically there are three
identifiable variations of abnormal pigmentation, namely hypopigmentation, hyperpigmentation and mixed pigmentation (Tyack et al. 1997; Van Zuijlen 2002).

Hypopigmented scars occur usually in deep injuries where all adnexal elements have been destroyed, and the only available source for melanocytes will be the wound edges (Chadwick et al. 2012). Hypopigmentation is aesthetically undesirable and the lack of melanin leaves hypopigmented scars without protection against ultraviolet radiation (UVR) leading to possible sunburn and malignancies (Chadwick et al. 2012).

Hyperpigmentation has commonly been attributed to effects of the inflammatory response during wound repair. The mechanisms involved in this process are not fully understood and may involve activation of melanocytes by inflammatory cell mediators released by injured skin. Post-burn injury hyperpigmentation is common after skin grafting (Chadwick et al. 2012).

Mixed pigmentation is considered to be a result of internal bleeding within the scar tissue, commonly due to pressure therapy or infection of scar tissue which leads to a pattern of mixed or mottled pigmentation which can potentially last many years (Tyack et al. 1997).

1.3.3.2 Vascularity

Angiogenesis is responsible for revascularisation of the newly formed granulation tissue. In hypertrophic scars, both the papillary dermis and the reticular dermis contain a greater number of blood vessels which have the tendency to be more dilated when compared to blood vessels in normal skin (Amadeu et al. 2003). There is also an increased flow of blood seen in the normal skin closest to the injury site due to vasodilation. These factors taken together aggregates the vascular response in the scar tissue (Church et al. 2006; Heng 2011). The vascular response in conjunction with the epidermal melanin content is likely an indication of the changes in the scar colour (Tyack et al. 1997).

The extensively vascularised dermis with dilated vessels contribute to the appearance of the hypertrophic scar (Leung et al. 1989). The red/purple raised hypertrophic scar in the majority of cases goes through a period of overgrowth before maturing to a paler flatter scar containing fewer and less dilated vessels (Hambleton et al. 1992). Clinically, the process of scar maturation is observed by morphological changes and clinical symptoms such as scar colour transition (refer to Table 1-2) from purple to pink (Leung et al. 1989).
Table 1-2: Clinically observable changes in hypertrophic scars during maturation

<table>
<thead>
<tr>
<th>Colour</th>
<th>Consistency</th>
<th>Thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep purple</td>
<td>Very hard</td>
<td>Very thick</td>
</tr>
<tr>
<td>Purple</td>
<td>Hard</td>
<td>Thick</td>
</tr>
<tr>
<td>Red</td>
<td>Firm</td>
<td>Moderately thick</td>
</tr>
<tr>
<td>Pink</td>
<td>Soft</td>
<td>Thin</td>
</tr>
<tr>
<td>Slightly pink</td>
<td>Very soft</td>
<td>Very thin</td>
</tr>
</tbody>
</table>


1.3.3.3 Fibrosis

There is divergence from the normal wound repair process during the proliferation phase with inflammatory cells and fibroblasts contributing to an excessive production of extracellular matrix and disorganisation of the matrix structure. This matrix consists of densely-packed parallel collagen fibres in contrast to the normal basket-weave collagen fibre lattice seen in the uninjured skin and leads to the abnormal thickness and elastic properties of scar tissue (Van Zuijlen 2002). Similarly, during the final phase of wound repair there is an imbalance in the synthesis and breakdown of collagen fibres resulting in thicker dermal layer (Van Zuijlen 2002). This is clinically observed as elevated thickness of the scar ranging from very thick to thin (refer to Table 1-2) with compromised elastic properties (ranging from very hard to very soft) leading to lower functional ability when compared to normal skin (Leung et al. 1989).

1.4 Impact of scarring after burn injury

Patients with post burn scarring face a range of long term impacts due to their scars. These impacts include physical, psychosocial and cosmetic related outcomes. Most of these impacts have the potential to impact on the quality of life of burns survivors (Lawrence et al. 2012; Durani et al. 2009; Gangemi et al. 2008).

1.4.1 Cosmetic impacts

The appearance of the skin is pivotal for an individual's identity and self-esteem and may impact on inter-personal relations (Pham et al. 2007). Changes in the physical appearance of burn survivors due to scarring will have implications for social adjustment and continuity in their social relationships (Lawrence et al. 2004). Living with scars to both visible and non-visible areas of the body is challenging (Lawrence et al. 2004; Durani et al. 2009). Over 40% of burn survivors seek professional support up to two years post-burn injury to deal with community
integration. Approximately half of the children who suffer a burn injury struggle to re-integrate socially due to the lack of adaptive skills (Blakeney et al. 2007).

After a burn injury the perceptions of patients of their cosmetic outcome may affect their mental state and this may then impact on their acceptance of their ‘new cosmetic appearance’ (Esselman 2007). Many burn survivors experience anxiety over their cosmetic appearance and seek treatment to minimise the scars and are further distressed to realise that the scars will not disappear even after repeated treatment (Brown et al. 2008). Andrade and Semple (2006) demonstrate a significant association between the patient’s over-all satisfaction of the surgical procedure and the post-surgical cosmetic outcome.

1.4.2 Functional impacts

The impact on the functional aspects of life after a burn injury is well documented (Durani et al. 2009; Gangemi et al. 2008; Esselman 2007). Increased catabolism among severe burns patients leads to low body mass causing weakness, low functional capability and fatigue. These factors have serious implications for returning to work and a normal life after a burn injury with subsequent impacts on the quality of life (Esselman 2007). In addition to scarring, post-burn symptoms such as chronic pain and pruritus may also impact on the daily living of burn survivors by reducing their general mobility and self-care, ability to fall asleep and stay asleep, wear certain type of clothes and participate in leisure activities (Bloemen et al. 2009; Brown et al. 2008). These difficulties may affect mood, with subsequent complications such as depression and anxiety (Brown et al. 2008).

Contractures are a common complication of burn injury and lead to the reduction of movements in limbs and joints and may cause deformities (Esselman 2007; Woo & Seul 2000). Deep burns to the face may heal with formation of scars and contracture of the perioral tissues (microstomia) leading to reduction in oral movements. Microstomia restricts the ability to smile, impedes speech and limits the movement of the lower jaw (Maragakis & Garcia-Tempone 1998). Eyelid or lip ectropion and nasal stenosis are some of the other devastating contractures that may occur with facial burns (Satish & Kathju 2010).
1.4.3 Psychosocial impacts

Depression, anxiety, sleep disturbances and post-traumatic stress are some of the conditions that occur frequently within the burn population (Esselman 2007). The relationship between post-burn appearance due to scarring and body esteem and the belief of some burn survivors that their appearance is directly influenced by the state of their scars has been reported by Lawrence and others (2006). A systematic review published in 2012 (Lawrence et al. 2012) highlighted the absence of published studies that examine the link between scarring and appearance as judged by general public. The authors describe how burn survivors believed their appearance to be directly influenced by state of their scars. Furthermore, the authors indicate that to date there are no studies that have examined the link between scarring and appearance judged by general public (Lawrence et al. 2012). To fully establish the psychosocial impact of scarring, there needs to be a better method of assessing scars and appropriate studies that measure the link between scarring and the perceptions of the burn survivor and general public of this condition (Lawrence et al. 2012; Blakeney et al. 2007). All the published data on this topic are based on proxy variables for scar severity with most studies examining the link between TBSA and psychosocial outcome (Lawrence et al. 2012).

1.5 Assessment of post-burn scarring

Post-burn scarring is a key issue for both clinicians and researchers. Assessment of the scar ‘phenotype’ is essential for diagnosis, monitoring, evaluation and therapeutic management of scar pathology (Bayat et al. 2003).

1.5.1 Context of scar assessment

Scar is important from both the clinical and research perspectives. From a clinical aspect, assessment of burn scars allows diagnosis and monitoring of scarring in a patient over time, enabling evaluation of treatment efficacy and the altering of treatment(s) according to scar progression (Roques & Teot 2007). From a research perspective, scar assessments permit the development of better scar outcome measurements, investigation of factors that may affect scar outcomes and the development of clinical interventions that may reduce scarring (Roques & Teot 2007).

It is envisaged that if an objective and reliable scar assessment method was available, such a method would offer the ability to compare scar outcomes not only of individual patients over time but between patients as well (Roques & Teot 2007).
1.5.2 Current scar assessment methods

Current scar assessment methods used for both research and clinical settings can be broadly categorised as subjective (observer dependent) and objective (Fearmonti et al. 2010). Subjective scar assessments are simple paper based tools conducted by clinicians or observers and patients providing a pure subjective description of the scar. Objective scar assessments are conducted using a technical device to measure the physical attributes of scars in a more reproducible way. By using a technical device the need for the person conducting the scar assessment to make a decision/interpretation of the scar measurement is eliminated (Brusselaers et al. 2010b; Fearmonti et al. 2010).

There is no universally accepted ‘gold standard’ method of assessing scars (Idriss & Maibach 2009). Furthermore, most current scar assessment methods fail to provide functional data on the progress of scarring (Wood et al. 1996).

1.5.3 Important scar assessment parameters

Scar assessment measures clinically discernible scar parameters that set the scar ‘phenotype’ apart from normal skin. Abnormal pigmentation and vascularity produce changes in skin colour. Excessive fibrosis produces changes in skin thickness, skin pliability and surface topography. Damage to nerves causes symptoms of pain and itch. Other clinically discernible features of the scar are its area of the skin surface and border. Burn scar assessment using subjective scar assessment scales can measure a range of these parameters.

1.6 Scar assessment scales

Smith and others (1988) were the first to demonstrate the possibility of measuring cosmetic disfigurement reliably by rating colour photographs of scars and scoring on five scar parameters (irregularity, thickness, colour, cosmetic disfigurement and overall disfigurement) utilising four scar observers (Idriss & Maibach 2009; Verhaegen 2011). Since then there have been many scar assessment scales developed (refer to Table 1-3) and these scales demonstrate varying levels of reliability. No ideal scar assessment scale has been proposed that is suitable for all scar assessment purposes (Verhaegen 2011).

Current scar assessment scales are subjective in nature and provide a measure of the scar by the rater (clinician/observer/patient) performing the assessment (Brusselaers et al. 2010a). Some semi-quantitative scales have also been designed incorporating objective components such as histology (Fearmonti et al. 2010). There are many published studies and
reviews (Brusselaers et al. 2010a; Idriss & Maibach 2009; Roques & Teot 2007; Bloemen et al. 2011b; Tyack et al. 2012) of current scar assessment scales. Table 1-4 summarises the reliability and methods of the scar assessment scales used both in the research and the clinical settings.

The Vancouver Scar Scale (VSS) developed by Sullivan and others (1990), assesses four scar parameters (pigmentation, vascularity, pliability and height) and was the first reliability assessed burn scar assessment scale (Sullivan et al. 1990; Idriss & Maibach 2009; Fearmonti et al. 2010; Roques & Teot 2007). To date, the VSS is the most widely used scar assessment scale within the clinical setting (Lye et al. 2006). The VSS focuses on the clinician’s estimation of the scar alone and lacks patient perceptions of the scar. Sullivan and others (1990) proposed that adding patient perceptions of ‘pain’ and ‘itching’ to the VSS would address this limitation. The VSS shows ‘low to moderate’ inter-rater reliability and ‘low’ internal consistency when three raters are used (Sullivan et al. 1990; Draaijers et al. 2004c).

Published literature (Fearmonti et al. 2010; Idriss & Maibach 2009; Roques & Teot 2007) describes many limitations of the VSS in particular the low inter-rater reliability, difficulty in assessing pigmentation due to the confounding influence of vascularity, the lack of inclusion of patient perceptions in the scale and the difficulty of locating the test site for follow-up assessment over time. To address some of these limitations, many subsequent authors (Baryza & Baryza 1995; Nedelec et al. 2000; Draaijers et al. 2004c; Schwanholt et al. 1994; Forbes-Duchart et al. 2007; Fraulin et al. 1996) have made improvements to the original VSS and developed modified versions of the VSS (mVSS). Schwanholt and others (1994) modified the VSS by omitting assessment of the pigmentation component to achieve higher reliability. Baryza and Baryza (1995) introduced a plexiglass tool to improve the inter-rater reliability of the pigmentation and height measurements. The plexiglass tool is used to blanch the area to reduce the vascular component when measuring pigmentation and transparent ruler markings on the plexiglass tool to measure the scar height. The authors also added a new pigmentation category (mixed pigmentation) to achieve greater sensitivity and reported increased inter-rater reliability for all components other than height (Baryza & Baryza 1995). Nedelec and others (2000) modified the VSS by rescaling the parameters of pigmentation and height and including patient perceptions of pain and itching scored on a Visual Analogue Scale (VAS). Forbes-Duchart and others (2007) modified the VSS by adding a colour scale to increase the reliability of the vascularity scale among their indigenous population when developing their mVSS. Despite these mVSS modifications these scales remain subjective with ‘low to moderate’ inter-rater reliability.

The Seattle Scar Scale (SSS) was introduced in 1997 (Yeong et al. 1997) and is a photo-based scale with high inter-rater reliability (ICC range 0.85-0.97); however, inter-rater reliability was based on the average scores of eight raters and not pair-wise testing. Masters and others (2005)
modified the SSS by including a localisation technique to ensure the ability to assess a similar scar area over time when utilising the SSS. The modified SSS by Masters and others (2005) is the method-matching assessment of scars and photographs (MAPS). Despite this modification MAPS remains more costly and time consuming than the VSS.

Beausang and others (1998) proposed the Manchester Scar Scale (MSS) as a more ‘complete’ quantitative scar assessment scale. The MSS consists of three separate assessment components: clinical, photographic (VAS) and histological (numerical scale). Clinically the MSS assessed pigmentation and vascularity combined and named this ‘colour mis-match’. The inter-rater reliability was reported as high ($r = 0.87$); however, this is an average measure for 10 raters and the inter-rater reliability for individual rater pairs was not reported.

In 1998 the Hamilton Scar Scale was proposed by Crowe and others (1998). This was a photo-based scale which reported ‘excellent’ inter-rater reliability (refer to Table 1-4). The authors indicate that this scale has ‘excellent’ reliability when used by both novice and experienced scar raters.

The Patient and Observer Scar Assessment Scale (POSAS) proposed by Draaijers and others (Draaijers et al. 2004c) consists of two six-item scales: an observer scale and a patient scale scored on a 10 point rating scale. In addition to these two scales an overall opinion of the scar is sought from both the patient and observer. The POSAS measures scar colour (pigmentation and vascularity) pliability/stiffness, thickness, surface irregularity, pain and itching. The POSAS is reported to have higher inter-rater reliability (ICC 0.87-0.95) compared to the VSS (Draaijers et al. 2004c). Another advantage of the POSAS is the inclusion of the patient perception of the scar which is clearly lacking in the VSS (Brusselaers et al. 2010a). The psychometric properties of the POSAS have been tested using Rasch analysis, and the results show that the POSAS has adequate fit to the Rasch model for all the scale items other than surface area. To date, POSAS is the only scar assessment scale to have Rasch analysis conducted (Van der Wal et al. 2012a).

The Stony Brook Scar Assessment Scale (SBSES) has been proposed by Singer and others (2007) to evaluate linear scars by scoring for the presence or the absence of several scar parameters (refer to Table 1-3). With ‘good’ reliability this scale is more suited to assess scars in the short-term and has not been tested on burn scars.
### Table 1-3: Published studies of scar assessment scales

<table>
<thead>
<tr>
<th>Scar scale Author</th>
<th>Assessment method</th>
<th>Reliability results</th>
<th>Strengths of scale as discussed by authors and in reviews</th>
<th>Limitations of scale as discussed by authors and in reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vancouver Scar Scale (VSS) Sullivan et al. (1990)</td>
<td>Study setting: clinical (paediatric and adult) No. of subjects = 73 No. of scars = not indicated No. of raters = 3 Scale: clinical subjective numerical scale Area assessed: 1&quot; x 1&quot; area of scar Components assessed: • Pigmentation • Vascularity • Pliability • Height</td>
<td>Inter-rater reliability (Cohen’s k): Overall = 0.5 Cohen’s k (for components) Pigmentation = 0.3-0.5 Vascularity = 0.1-0.4 Pliability = 0.2-0.4 Height = 0.1-0.5</td>
<td>Designed specifically for burn scars The first burn scar scale that was reliability tested Paper based/easy to use/ no cost Most widely used in literature in burns at present</td>
<td>Low inter-rater variability Lacks patient perception of the scar Difficult use in large scars with non-uniform pigmentation patterns Difficult to distinguish colour within the vascularity subscale Accurate height measures in millimetres is problematic Difficult to distinguish between contracture and pliability Does not locate the test site within scar for follow up testing.</td>
</tr>
<tr>
<td>Modified VSS Schwanholt et al. (1994)</td>
<td>Study setting: clinical (paediatric) No. of subjects = 63 No. of scars = scars not indicated No of raters = 2 Scale: clinical subjective numerical scale Area assessed: 1&quot; x 1&quot; area of scar Modification from the original VSS = pigmentation not assessed, compliance of pressure garment use included Components assessed: • Vascularity • Pliability • Height • Compliance (pressure garment)</td>
<td>Inter-rater reliability data not indicated (study aimed to compare scar maturation in different age groups)</td>
<td>Not discussed</td>
<td>Not discussed</td>
</tr>
<tr>
<td>Scar scale Author</td>
<td>Assessment method</td>
<td>Reliability results</td>
<td>Strengths of scale as discussed by authors and in reviews</td>
<td>Limitations of scale as discussed by authors and in reviews</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>----------------------------------------------------------</td>
<td>----------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Modified VSS      | Study setting: clinical  
No. of subjects = not indicated  
No. of raters = not indicated  
Scale: clinical subjective numerical scale with a plexiglass tool  
Modification from the original VSS = inclusion of the plexiglass tool for measuring pigmentation and height  
Area assessed: not indicated  
Components assessed:  
- Pigmentation  
- Vascularity  
- Pliability  
- Height | Inter-rater reliability:  
ICC (overall) = 0.81  
Cohen’s k (for components)  
Pigmentation = 0.61  
Vascularity = 0.73  
Pliability = 0.71  
Height = 0.56 | Higher reliability than VSS  
Designed specifically for burn scars | Higher cost than VSS |
No. of subjects = not indicated  
No. of raters = not indicated  
Scale: clinical subjective numerical scale  
Modification from the original VSS = measurement of facial symmetry  
Area assessed: not indicated  
Components assessed:  
- Pigmentation  
- Vascularity  
- Pliability  
- Height  
- Facial symmetry | Inter-rater reliability data not reported (study aimed to compare outcomes of two interventions) | Not discussed | Not discussed |
| Modified VSS      | Study setting: clinical (paediatric and adult)  
No. of subjects = 48  
No. of scars = 48  
No. of raters = not indicated  
Scale: clinical subjective numerical scale  
Area assessed: not indicated  
Modification from the original VSS = measurement of facial symmetry  
Components assessed:  
- Pigmentation  
- Vascularity  
- Pliability  
- Height  
- Facial symmetry | Inter-rater reliability data not reported (study aimed to compare outcomes of two interventions) | Not discussed | Not discussed |
<table>
<thead>
<tr>
<th>Scar scale</th>
<th>Assessment method</th>
<th>Reliability results</th>
<th>Strengths of scale as discussed by authors and in reviews</th>
<th>Limitations of scale as discussed by authors and in reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seattle Scar Scale (SSS)</td>
<td>Study setting: clinical</td>
<td>Inter-rater reliability:</td>
<td>Higher inter-rater reliability than the VSS</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Yeong et al. (1997)</td>
<td>No. of subjects = 10</td>
<td>ICC (for components)</td>
<td>Able to assess scars without patient being present</td>
<td>Higher cost compared to the VSS</td>
</tr>
<tr>
<td></td>
<td>No. of scars = not indicated</td>
<td>Surface = 0.97</td>
<td></td>
<td>Difficulty in assessing thickness and surface from photographs</td>
</tr>
<tr>
<td></td>
<td>No. of raters = 8</td>
<td>Thickness = 0.93</td>
<td></td>
<td>Difficulty in comparing non-uniform scars to the standardised photographs</td>
</tr>
<tr>
<td></td>
<td>Scale: photo based scale</td>
<td>Border = 0.95</td>
<td></td>
<td>Scoring of hypopigmentation may not be applicable for all skin types</td>
</tr>
<tr>
<td></td>
<td>Area assessed: ‘whole scar’</td>
<td>Colour = 0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Components assessed:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Surface</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Border</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Thickness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Colour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study setting: clinical (adults)</td>
<td>Inter-rater reliability:</td>
<td>Provides a comprehensive scar assessment with both clinical and histological data</td>
<td>The need for three assessments on each subject not viable clinically</td>
</tr>
<tr>
<td>Manchester Scar Scale (MSS)</td>
<td>No. of subjects = 55</td>
<td>ICC (for components)</td>
<td>Applicable to a wide variety of scars (surgical, hypertrophic and keloid)</td>
<td>Time (over 20 min)</td>
</tr>
<tr>
<td>Beausang et al. (1998)</td>
<td>No. of scars = 14</td>
<td>Surface = 0.97</td>
<td>Better suited for assessing linear scars</td>
<td>Costly</td>
</tr>
<tr>
<td></td>
<td>No. of raters = 10</td>
<td>Thickness = 0.93</td>
<td></td>
<td>Does not assess vascularity separately</td>
</tr>
<tr>
<td></td>
<td>Scale: clinical, photographic (VAS) and histologic (numerical scale)</td>
<td>Border = 0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Area assessed: ‘whole scar’</td>
<td>Colour = 0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Components assessed:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Clinically and photographically</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Colour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Contour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Texture and distortion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Matt vs. shiny</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Histologically</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Collagen orientation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Collagen density</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Collagen maturity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scar scale</td>
<td>Assessment method</td>
<td>Reliability results</td>
<td>Strengths of scale as discussed by authors and in reviews</td>
<td>Limitations of scale as discussed by authors and in reviews</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>------------------------------------------</td>
<td>------------------------------------------------------------</td>
<td>-------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Hamilton Scar Scale | Study setting: clinical (adults)  
No. of subjects = 10  
No. of scars = 10  
No. of raters = 4  
Scale: photo based scale  
Area assessed: not indicated  
Components assessed:  
- Surface irregularity  
- height  
- Colour  
- Vascularity | Inter-rater reliability:  
k_w (for components)  
Irregularity = 0.76  
Height = 0.79  
Colour = 0.90  
Vascularity = 0.66  
Test-retest reliability:  
k_w (for components)  
Irregularity = 0.80  
Height = 0.84  
Colour = 0.85  
Vascularity = 0.78 | Highly reliable for use by novice and experienced therapists  
Requires minimal training | Does not include patient perception |
| Modified VSS | Study setting: clinical  
No. of subjects = 15  
No. of scars = 18  
No. of raters = 3  
Scale: clinical subjective numerical and visual analogue scale (VAS) scale  
Area assessed: not indicated  
Modification from the original VSS = inclusion of pain and itching  
Components assessed:  
- Pigmentation  
- Vascularity  
- Pliability  
- Height  
- Pain and itching | Inter-rater reliability:  
ICC = 0.51-0.57 (overall)  
Cohen’s k (for components)  
Pigmentation = 0.14-0.24  
Vascularity = 0.14-0.25  
Pliability = 0.19-0.22  
Height = 0.07-0.25 | Training in the use of this scale is required  
Documentation of additional pertinent information | Not discussed |
<table>
<thead>
<tr>
<th>Scar scale</th>
<th>Assessment method</th>
<th>Reliability results</th>
<th>Strengths of scale as discussed by authors and in reviews</th>
<th>Limitations of scale as discussed by authors and in reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified VSS</td>
<td>Study setting: clinical (adult)</td>
<td>Inter-rater reliability:</td>
<td>Not discussed</td>
<td>Lower internal consistency than the POSAS</td>
</tr>
<tr>
<td>Draaijers et al. (2004c)</td>
<td>No. of subjects = 20</td>
<td>ICC = 0.90 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No. of scars = 49</td>
<td>ICC = 0.69 (single rater pair)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No of raters = 4</td>
<td>Internal consistency:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scale: clinical subjective numerical scale</td>
<td>Cronbach’s alpha = 0.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Area assessed: not indicated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modification from the original VSS = inclusion of mixed pigmentation and modification to height scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Components assessed:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Pigmentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Vascularity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Pliability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Height</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient and Observer Assessment Scale (POSAS)</td>
<td>Study setting: clinical (adults)</td>
<td>Inter-rater reliability:</td>
<td>Good inter-rater reliability and internal consistency compared to the VSS</td>
<td>Difficult to use in young children</td>
</tr>
<tr>
<td>Draaijers et al. (2004c)</td>
<td>No. of subjects = 20</td>
<td>ICC = 0.92 (average)</td>
<td>Reliability tested for use in burn scars</td>
<td>Test-retest reliability and intra-observer reliability have not been established for burn scars</td>
</tr>
<tr>
<td></td>
<td>No. of scars = 49</td>
<td>ICC = 0.73 (pairwise)</td>
<td>Captures both the patient and the clinician’s perception of the scar</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No of raters = 4</td>
<td>Pigmentation = 0.77-0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scale: clinical subjective, two numeric scales (observer / patient )</td>
<td>Vascularity = 0.87-0.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Area assessed: not indicated</td>
<td>Internal consistency:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Components assessed:</td>
<td>Cronbach’s alpha = 0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Colour (pigmentation/vascularity)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Pliability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Thickness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Relief</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Pain and itching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Irregularity and stiffness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scar scale</td>
<td>Assessment method</td>
<td>Reliability results</td>
<td>Strengths of scale as discussed by authors and in reviews</td>
<td>Limitations of scale as discussed by authors and in reviews</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Modified SSS</td>
<td>Study setting: clinical</td>
<td>Inter-rater reliability:</td>
<td>Can be used for burn scars</td>
<td>Complex and time consuming</td>
</tr>
<tr>
<td>Modified SSS</td>
<td>No. of subjects = 5</td>
<td>ICC (for components)</td>
<td>Ability to assess similar area over time due to localisation technique</td>
<td>Higher cost compared to the VSS</td>
</tr>
<tr>
<td>Matching Assessment of Scars and Photographs (MAPS)</td>
<td>No. of scars = 32</td>
<td>Thickness = 0.62-0.82, Border height = 0.74-0.78, Colour = 0.64-0.71, Surface = 0.25-0.38</td>
<td>Good for long-term follow up</td>
<td>Raters require training in the use of the MAPS</td>
</tr>
<tr>
<td>Masters et al. (2005)</td>
<td>Scale: photo based scale, with a numeric scale and a body map</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Area assessed: ‘whole scar’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Components assessed:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Thickness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Border height</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Colour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Surface</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified VSS</td>
<td>Study setting: clinical (Aboriginal and Caucasian)</td>
<td>Inter-rater reliability:</td>
<td>Pictorial colour scale appears to be a useful feature of this mVSS</td>
<td>Pigmentation – poor reliability</td>
</tr>
<tr>
<td>Forbes-Duchart et al. (2007)</td>
<td>No. of subjects = 14</td>
<td>ICC = 0.76-0.84 (overall)</td>
<td></td>
<td>Vascularity – poor reliability in Aboriginal</td>
</tr>
<tr>
<td></td>
<td>No of scars = 32</td>
<td>k_w (for components)</td>
<td></td>
<td>k_w values very low overall – may be due to small sample size</td>
</tr>
<tr>
<td></td>
<td>No. of raters = 3</td>
<td>Pigmentation = 0.21-0.16, Vascularity = 0.04-0.25, Pliability = 0.12-0.38, Height = 0.54-0.58</td>
<td>Additional colour scales may be required (not ethnicity based)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scale: clinical subjective numerical scale with a plexiglass tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Area assessed: not indicated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modification from the original VSS = inclusion of a colour chart to assess pigmentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Components assessed:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Pigmentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Vascularity (colour scale)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Pliability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Height</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scar scale Author</td>
<td>Assessment method</td>
<td>Reliability results</td>
<td>Strengths of scale as discussed by authors and in reviews</td>
<td>Limitations of scale as discussed by authors and in reviews</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>The Stony Brook Scar Assessment Scale (SBSES)</td>
<td>Study setting: Clinical</td>
<td>Inter-rater reliability:</td>
<td>Developed to assess short-term scar appearance</td>
<td>Proposed for use in research only recently</td>
</tr>
<tr>
<td>Singer et al. (2007)</td>
<td>No. of subjects = 50</td>
<td>Spearman rho (overall) = 0.73-0.85</td>
<td>Developed to assess linear scars</td>
<td>Not tested in burns scars</td>
</tr>
<tr>
<td></td>
<td>No. of scars = 50</td>
<td>Spearman rho (for components</td>
<td></td>
<td>Does not include patient perception</td>
</tr>
<tr>
<td></td>
<td>No. of raters = 3</td>
<td>Width = 0.62-0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scale: Photo based scale with a VAS</td>
<td>Height = 0.64-0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Area assessed: not indicated</td>
<td>Colour = 0.67-0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Components assessed:</td>
<td>Hatch marks = 0.79-0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Scar width and height</td>
<td>Overall appearance = 0.41-0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Colour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Hatch marks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Overall appearance</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

24
1.7 Scar assessment devices

At present there are many devices used within the research and clinical setting to objectively measure one or more characteristics of a scar (Fearmonti et al. 2010; Roques & Teot 2007). To date there is no single device that is capable of measuring all the important characteristics of a scar in a reliable and reproducible manner (Brusselaers et al. 2010b; Roques & Teot 2007; Fearmonti et al. 2010). The advantages of scar assessment devices are that they measure a physical scar parameter in a quantitative manner providing an objective and reproducible measurement eliminating rater judgement and potential bias from the scar measurement, which is often seen in the paper-based subjective scar scales (Brusselaers et al. 2010b).

Scar assessment devices have a few disadvantages over the scar assessment scales. Firstly, the devices often measure only a single scar parameter. Secondly, the devices are unable to measure physiological criteria such as pain and itching which have been found to have profound effects on the quality of life of the burns patient (Brusselaers et al. 2010b). Thirdly, devices are costly and require training, and finally, other patient perceptions (e.g. dissatisfaction with scar, embarrassment) are not taken into account when a device is used.

The scar assessment devices available within the research and clinical settings at present can be broadly divided into three distinct groups based on the scar parameter measured i.e. colour (pigmentation/vascularity), pliability/elasticity and height/thickness (Brusselaers et al. 2010b). Within these three groups, the devices are further subdivided into sub-groups based on the principle of measurement utilised for measurement (refer to Tables 1-4 to 1-6).

1.7.1 Colour evaluation devices

The devices described in this section measure the colour of a scar in a quantitative manner and provide an objective measurement. The assessment of scar colour is conducted by a variety of physical principles including reflection/absorption of light, laser doppler imaging and computer analysis of colour. Table 1-4 presents a summary of devices described in published literature which measure scar colour. Despite the availability of a variety of different principles for colour assessment, the most frequently used principle of measurement is spectrophotometry (Van der Wal 2013). There are two sub-types of spectrophotometry: tristimulus reflectance colorimetry (used in devices such as Chromameter CR-200/300 and Labscan XE) and narrow band spectrophotometry (used in devices such as the DermaSpectrometer and Mexameter).

The Chromameter CR-200/200 and the Labscan XE both have ‘excellent’ inter-rater reliability on normal skin and ‘good’ reliability on scar tissue (Van der Kerkhove et al. 2005; Li-Tsang et al. 2003). Validation of the Chromameter CR-200/200 was undertaken using the POSAS as the
reference standard while the Labscan XE validation used the VSS as the reference standard. Both devices showed low fit for vascularity and pigmentation measurements to respective reference standards used (Li-Tsang et al. 2003; Draaijers et al. 2004b). The Chromameter CR-200/200 is currently not commercially available, and due to the high cost and lack of portability the Labscan XE has limited clinical usefulness (Brusselaers et al. 2010b).

Narrow-band spectrophotometry (DermaSpectrometer and Mexameter) also measures pigmentation (melanin) and vascularity (Haemoglobin [Hb]) with measurements expressed as an index: for pigmentation (melanin index [MI]) and for vascularity (erythema index [EI]) (Van Zuijlen et al. 2002). The DermaSpectrometer and the Mexameter both demonstrate ‘excellent’ inter-rater reliability on normal skin and scar tissue (Draaijers et al. 2004b; Clarys et al. 2000; Nedelec et al. 2008). The validation of the DermaSpectrometer was undertaken using the POSAS, with the VSS used as the reference standard for the Mexameter. Each device demonstrated a ‘good’ fit for the vascularity parameter and a ‘low’ for the pigmentation parameter (Nedelec et al. 2008). The DermaSpectrometer is currently not available as the model has been superseded (Van der Wal 2013).

The other modalities of colour assessment described in Table 1-5 mentioned in the published literature are Laser doppler imaging (Brusselaers et al. 2010b), computer analysis of colour (Davey et al. 1999) and optical coherence tomography (Liew et al. 2013). These methods lack reliability and validity information making them less suitable for clinical application in the burn arena.

1.7.2 Height/thickness evaluation devices

Subjective scar assessment scales rate scar height by palpation and visualisation of the scar in comparison to the adjacent normal skin, or use of simple ruler. Therefore scales are a crude form of measurement and assess only the palpable and visible portion (height) of the scar. This measurement does not reflect the true thickness of scar as a part of the scar lies within the dermis and is not visible to the human eye (Brusselaers et al. 2010b). The current ‘gold standard’ for measuring thickness is a biopsy. A biopsy is used rarely due to its invasive nature. Therefore to measure the total thickness of a scar, a non-invasive, reliable and easy to use device is needed (Van Zuijlen et al. 2002). Table 1-5 describes the range of devices (categorised by the principle of measurement) that are available for measuring scar height/thickness. The three dimensional (3D) methods suggested by Sawada (1994) are not yet clinically viable while the magnetic resonance imaging (MRI) (Muthupillai et al. 1995) remains not portable and too expensive for use in clinical scar assessment (Brusselaers et al. 2010b). The ultrasound methods are the most suitable and widely used (Brusselaers et al. 2010b; Fearmonti et al. 2010). Of the
ultrasound-based devices, the Tissue Ultrasound Palpation (TUPS) (Lau et al. 2005) and DermaScan C (Nedelec et al. 2008) show the most potential for clinical scar assessment with similar reliability and validity statistics (Brusselaers et al. 2010b).

The remaining modality, the Vivid 900 3D digitiser (Taylor et al. 2007), has not yet been tested on burn scars and lacks reliability data (Van der Wal 2013). However this type of device has potential to generate important spatial and topographical scar information in the future.

1.7.3 Pliability/elasticity evaluation devices

The devices that are used to evaluate elasticity of the skin originate from different fields and have been adapted for scar evaluation. Brusselaers and others (Brusselaers et al. 2010b) indicate that most of these elasticity measuring devices have originated from fields such as dermatology, ophthalmology, and other industrial applications.

Elastometers or elasticity measuring devices measure the elastic properties of the skin using a range of different biomechanical force such as suction, pressure, torsion and tension (Brusselaers et al. 2010b; Van Zuijlen et al. 2002). Table 1-6 describes the pliability/elasticity measuring devices described in literature categorised by the biomechanical force used for the measurement of elasticity.

Among the devices that use the suction principle, the Cutometer demonstrates ‘good’ inter-rater reliability for both normal skin and scar tissue but has low validity with the POSAS (Nedelec et al. 2008; Draaijers et al. 2004a). The DermaFlex has now been replaced with the DermaLab (Cortex Technologies 2011). The DermaLab is the device that Cortex Technologies developed prior to the DermaLab Combo and had the capacity to measure only two out of the four significant scar parameters (elasticity and trans-epidermal water loss). The DermaLab measures of elasticity and trans-epidermal water loss demonstrated ‘excellent’ inter-rater reliability; validation (only elasticity) using the DermaFlex as reference standard demonstrated ‘moderate’ validity (Anthonissen et al. 2012; Pederson & Jemec 2006).

The tonometer is the most commonly used device among the group of devices that use the pressure principle for measuring elasticity. The tonometer has ‘excellent’ inter-rater reliability and ‘moderate’ validity with the pliability score of the VSS (Lye et al. 2006; Corica et al. 2006; Merkel et al. 2008; Katz et al. 1985).

The dermal torque meter (torsion as the elasticity measuring principle) also lacks reliability data (Murray & Wickett 1997; Boyce et al. 2000), and the Reviscometer (tension as the elasticity measuring principle) is reliable but requires validation (Verhaegen et al. 2010).
Therefore among the currently available elastometers, the Cutometer and the tonometer are the devices of choice due to their reliability. However, the tonometer has the limitation of not being suitable for measurement in locations with close proximity to bony structures and the possible discomfort to patients (Brusselaers et al. 2010b; Fearmonti et al. 2010).
Table 1-4: Summary of devices that measure scar colour categorised by the principle of measurement

<table>
<thead>
<tr>
<th>Device (Manufacturer) Reference</th>
<th>Assessed on</th>
<th>Inter-rater reliability (ICC)</th>
<th>Validity</th>
<th>Strengths / Limitations of device as discussed by authors and in the reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reflection/ absorption of light (spectrophotometry)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tristimulus reflectance colorimeters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromameter CR-300 (Minolta) Van den Kerckhove et al. (2005)</td>
<td>Normal skin</td>
<td>Normal skin:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burn scar</td>
<td>Pigmentation = 0.92 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vascularity = 0.97 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Burns scar:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pigmentation = 0.73-0.89 (pair-wise), 0.91 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vascularity = 0.75 (pair-wise), 0.92 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Validated with POSAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spearman rho:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pigmentation = 0.23-0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vascularity = 0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Strengths:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>More reliable than scar scales</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Immediate on-site assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limitations:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not available commercially</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labscan XE (Hunter Lab) Li-Tsang et al. (2003)</td>
<td>Normal skin</td>
<td>Normal skin:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burn scar</td>
<td>Pigmentation = 0.95-0.96 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vascularity = 0.92 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Burn scar:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pigmentation = 0.88-0.99 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vascularity = 0.50 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Validated with VSS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spearman rho:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pigmentation = 0.50-0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vascularity = 0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limitations:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not portable and expensive</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Narrow-band spectrophotometers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DermaSpectrometer (Cortex Technologies) Draaijers et al. (2004b) Clarys et al. (2000)</td>
<td>Burn scar</td>
<td>Burn scar:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pigmentation = 0.94 (pair-wise), 0.98 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vascularity = 0.72 (pair-wise), 0.91 (average)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Validated with POSAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spearman rho:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pigmentation = 0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vascularity = 0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Strengths:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Portable, easy to use</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limitations:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not available commercially</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device (Manufacturer)</td>
<td>Assessed on</td>
<td>Inter-rater reliability (ICC)</td>
<td>Validity</td>
<td>Strengths / Limitations of device as discussed by authors and in the reviews</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------</td>
<td>-------------------------------</td>
<td>----------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Narrow-band spectrophotometers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Mexameter (Khazaka Electronics) | Normal skin
Nedelec et al. (2008) | Normal skin:
• Pigmentation = 0.99 (average)
• Vascularity = 0.97 (average)
Burn scar:
• Pigmentation = 0.95-0.98 (average)
• Vascularity = 0.82-0.97 (average) | Validated with VSS
Spearman rho:
Pigmentation = 0.37-0.50
Vascularity = 0.52-0.65 | Strengths:
• Portable, easy to use |
| **Laser based methods** |             |                               |          |                                                                         |
| The Laser-Doppler Imaging (LDI) (Teledyne Brown Engineering) Brusselaers et al. (2010) | Normal skin
Burn scar | Not available | Not available | Strengths:
• Assesses vascular density
Limitation:
• Does not assess melanin
• Not portable and expensive |
| Optical coherence tomography (Thorlabs) Liew et al. (2013) | Normal skin
Burn scar | Not available | Not available | Strengths:
• Accurate vascularisation pattern
Limitation:
• Measures only Hb content |
| **Computer software and video/photography** |             |                               |          |                                                                         |
| Computer analysis of colour (Sony Hi-8 Handycam and 486 PC-style computer) Davy et al. (1999) | Normal skin
Burn scar | Not available | Not available | Strength:
• Assess without patient presence
Limitations:
• Reliability not tested
• Light conditions, camera settings influence readings |
<table>
<thead>
<tr>
<th>Device (Manufacturer)</th>
<th>Assessed on</th>
<th>Inter-rater reliability (ICC)</th>
<th>Validity</th>
<th>Strengths/Limitations of device as discussed by authors and in the reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3 Dimensional (3D) Methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Negative– positive moulds/ replicas | Burn scar | Not available | Not available | Strengths:  
- Accurate, inexpensive  
Limitations:  
- Not clinically viable |
| Sawada (1994) | | | | |
| Magnetic Resonance Imaging (MRI) (Thorlabs) | Normal skin | Not available | Not available | Strengths:  
- Non-contact/high resolution/ precision  
Limitations:  
- High cost, not portable |
| Muthupillai et al. (1995) | | | | |
| **Ultrasound Methods** | | | | |
| Tissue ultrasound palpation system (TUPS) (Biomedical Ultrasonic Solutions) | Burn scar | Scar= 0.84 (average) | Validated with VSS Spearman rho:  
Height = 0.34 (average) | Strengths:  
- Portable, reliable  
Limitations:  
- Small probe diameter (not reflective of the whole scar) |
| Lau et al. (2005) | | | | |
| DermaScan C (Cortex Technologies) | Burn scar | Normal = 0.85 (average)  
Scar = 0.89-0.91 (average) | Validated with VSS Spearman rho:  
Height = 0.41-0.50 (average) | Strengths:  
- Reliable accurate, portable  
Limitation: Not discussed |
| Nedelec et al. (2008) | | | | |
| Vivid 900 3D Digitiser (Minolta) | Surgical scar  
Keloid scar | Not available | Validated with MSS Pearson’s r:  
Height = 0.63 (average) | Strengths:  
- Non-contact, repeatable  
Limitations:  
- Not tested on burn scars |
<p>| Taylor et al. (2007) | | | | |</p>
<table>
<thead>
<tr>
<th>Device (Manufacturer)</th>
<th>Assessed on</th>
<th>Inter-rater reliability (ICC)</th>
<th>Validity</th>
<th>Strengths/Limitations of device as discussed by authors and in the reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suction mechanism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cutometer SEM 474</td>
<td>Normal skin</td>
<td>Normal skin: 0.75-0.85 (pair-wise)</td>
<td>Validated with POSAS</td>
<td>Strengths: Reliable, portable and user friendly</td>
</tr>
<tr>
<td>(Khazaka Electronics)</td>
<td>Burn scar</td>
<td>0.95 (average) Burn scar: 0.35-0.76 (rater pair) 0.68-0.93 (average)</td>
<td>Spearman rho: 0.29-0.53 (average)</td>
<td></td>
</tr>
<tr>
<td>Nedelec et al. (2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Draaijers et al. (2004a)</td>
<td>Normal skin</td>
<td>Not available</td>
<td>Validated with DermaLab®</td>
<td>Strengths: not discussed</td>
</tr>
<tr>
<td>Dermaflex®</td>
<td>Normal skin</td>
<td>Normal skin: 0.93 (average) Burn scar: 0.86 (average)</td>
<td>Spearman rho: 0.38-0.44 (average)</td>
<td>Limitations: Not available commercially</td>
</tr>
<tr>
<td>(Cortex Technologies)</td>
<td>Burn scar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedersen and Jemec, (2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DermaLab®</td>
<td>Normal skin</td>
<td>Normal skin: 0.82-0.91(average) Scleroderma: 0.95 (average) Burn scar: 0.91 (average)</td>
<td>Validated with VSS</td>
<td>Strengths: not discussed</td>
</tr>
<tr>
<td>(Cortex Technologies)</td>
<td>Scleroderma</td>
<td></td>
<td>Spearman rho: 0.44-0.46 (average)</td>
<td>Limitations: Bony structures in close proximity may interfere Some discomfort to patients</td>
</tr>
<tr>
<td>Anthonissen et al. (2012)</td>
<td>Burn scar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pressure mechanism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tonometer (generic term)</td>
<td>Normal skin</td>
<td>Normal skin: 0.82-0.91(average) Scleroderma: 0.95 (average) Burn scar: 0.91 (average)</td>
<td>Validated with VSS</td>
<td>Strengths: not discussed</td>
</tr>
<tr>
<td>Lye et al. (2006)</td>
<td>Scleroderma</td>
<td></td>
<td>Spearman rho: 0.44-0.46 (average)</td>
<td>Limitations: Bony structures in close proximity may interfere Some discomfort to patients</td>
</tr>
<tr>
<td>Corica et al. (2006)</td>
<td>Burn scar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merkel et al. (2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Katz et al. (1985)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device (Manufacturer) Reference</td>
<td>Assessed on</td>
<td>Inter-rater reliability (ICC)</td>
<td>Validity</td>
<td>Strengths/Limitations of device as discussed by authors and in the reviews</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
<td>-------------------------------</td>
<td>----------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Torsion mechanism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dermal Torque meter (Dia_Stron, UK) Murray and Wickett (1997) Boyce (2000)</td>
<td>Burn scar</td>
<td>Not available</td>
<td>Poor with Cutometer</td>
<td>Strengths: not discussed Limitations: • Not reliability tested</td>
</tr>
<tr>
<td><strong>Tension mechanism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviscometer® (Khazaka Electronics) Verhaegen et al. (2010)</td>
<td>Normal skin Burn scar</td>
<td>Normal skin: 0.66-0.86 (pair-wise) 0.79-0.93 (average) Burn scar: 0.75-0.94 (pair-wise) 0.86-0.96 (average)</td>
<td>Not available</td>
<td>Strengths: not discussed Limitations: • Age, BMI, room temperature, humidity may impact results</td>
</tr>
</tbody>
</table>
1.7.4 Multi-component scar testing devices

The DermaLab® (Cortex Technologies, Denmark) is an example of a multi-component device with capacity to measure two scar parameters, pliability and trans-epidermal water loss (TEWL). The DermaLab® was tested by Anthonissen and others (2012) in burn scars and was found to demonstrate high inter-rater reliability for pliability (ICC 0.86-0.93) and ‘good’ (ICC 0.78-0.93) inter-rater reliability for TEWL (Anthonissen et al. 2012). The DermaLab® elasticity measurements were validated using the Dermaflex® (Cortex Technologies, Denmark) as a reference standard and demonstrated ‘moderate’ (0.38-0.43) validity. However the validation was conducted on normal skin (Pederson et al. 2003).

The DermaLab Combo® (Cortex Technologies, Denmark) is a multi-component skin testing device developed to replace the DermaLab®. The DermaLab Combo® has the capacity to measure multiple scar parameters including pigmentation, vascularity, pliability, thickness, sebum, trans-epidermal water loss and moisture. Therefore, the DermaLab Combo® has potential to make objective measurements of burn scar parameters analogous to the VSS.

The DermaLab Combo® consists of a main unit with screen (dimensions of 23 x 23 x 24 cm) embedded with WINDOWS operating system, a storage capacity of 1 GB RAM and a 8.4 inch colour touch screen (800 x 600 pixels). The device uses LabView® based software and provides USB connectivity. Multiple separate probes are available based on proven technologies such as narrow-band spectrophotometry (melanin and haemoglobin), high-frequency ultrasound (thickness) and suction (pliability). All probes have separate channel connectors with separate tubes that are then attached to the main unit (Cortex Technologies 2011).

The DermaLab Combo® colour probe is used to measure pigmentation and vascularity. This probe is based on the principle of narrow-band reflectance spectrophotometry (550 nm ±30 nm and 660 nm ±60 nm for haemoglobin [erythema] and melanin [pigmentation] respectively). The colour probe has an optical focussing on 7 mm diameter target area with a clear front for accurate positioning and is illuminated by two angled white light emitting diode (LED) lights. The probe displays four readings (three individual measurements with their average) separately for erythema (vascularity) and melanin (pigmentation) using Commission Internationale de l’Eclairage (CIE) – luminance (L), red-green axis (a) and blue-yellow axis (b) [CIELab] values (Cortex Technologies 2011).

The elasticity probe of the DermaLab Combo® measures pliability. This probe is based on the principle of vertical suction applied on the surface of scar. The probe has a measuring aperture of 10 mm diameter and adheres to the skin by a double adhesive sticker. Elasticity measurements are expressed in terms of Young’s modulus in megaPascals [MPa] (Cortex Technologies 2011).
The ultrasound probe of the DermaLab Combo® measures the skin elasticity/thickness. Based on the principle of high frequency ultrasound (20 MHz) this probe has a resolution of 60 x 200 micrometres with a penetration capacity of 3.4 mm with a fully adjustable gain settling (±10 dB). The probe has a rotating transducer, scan length 17 mm, footprint 11 mm, and has the capacity to display actual and stored measurements side-by-side (Cortex Technologies 2011).

The DermaLab Combo® moisture probe is based on the principle of conductance and uses a single frequency and provides a read-out within the range of 0–9999 microSiemens. The average of eight measurements is displayed on the screen. The moisture probe consists of eight pins minimising moisture accumulation and has a spring loaded action that triggers/ stops measurement. The resolution is 1 microSiemens with an accuracy of 5% (Cortex Technologies 2011).

The DermaLab Combo® trans epidermal water loss (TEWL) probe uses the principle of diffusion gradient (range of 0–250 g/m²/h. resolution: 0.1 g/m²/h and accuracy: 5%). The probe consists of two combined humidity/temperature sensors in 10 mm cylindrical diffusion chamber and is complete with an auto-stop mechanism. The main unit displays a continuous TEWL (five second mean) real time curve and sensor display (Cortex Technologies 2011).

The DermaLab Combo® sebum probe collects material on a microporous polymer film and uses translucency characteristics of sebum as the measuring principle (range of 0–100.0 %, resolution: 0.1 % with accuracy of 5%) (Cortex Technologies 2011). In addition, the DermaLab Combo® is also equipped with a video scope that has a resolution: 1.3 megapixel with a magnification capacity of 10 (Cortex Technologies 2011).

In summary, the DermaLab Combo® uses established technologies, is portable, non-invasive and relatively inexpensive (AUS$16,000 with all probes). Whilst the DermaLab Combo® was primarily designed for the cosmetic field, there may be potential for its use in burns scar assessment.

1.8 Summary

The aims of the literature review were to examine the available scar assessment methods and to identify a potential objective device with capacity to measure important scar parameters. The DermaLab Combo® was identified as a potential multi-component device to be tested for assessing scar parameters analogous to those in the modified VSS used at the Royal Perth Hospital. Post-burn scars evolve continuously during the maturation process and the clinical signs associated with the maturation process can be measured with either scar assessment scales or objective devices. The scar scales are subjective with varying levels of reliability and
validity. While the VSS, the modified versions of the VSS (mVSS), the POSAS and the MSS assess multiple scar parameters, are easy to use and cost-effective, better methods are needed.

The majority of the available devices identified in the literature review had the capacity to measure only one scar parameter and demonstrated varying levels of reliability. Two multi-component skin testing devices developed by Cortex Technologies, Denmark, were identified: the DermaLab® and the DermaLab Combo®. The DermaLab® has capacity to measure pliability and trans-epidermal water loss with good to excellent reliability (Anthonissen et al. 2012) with ‘moderate’ validity demonstrated for pliability assessment of burn scars, using the DermaFlex® as reference standard (Anthonissen et al. 2012; Pederson & Jemec 2006). The DermaLab Combo® represents the next generation device, replacing the DermaLab®, and has the capacity to measure four scar assessment parameters relevant to burn scars: pigmentation, vascularity, pliability and thickness.

1.9 Project rationale and aims

Burn injury wounds remain one of the most complex acute wounds to treat. Despite the advances in burn care over the recent decades, scarring remains a problem for clinicians and burn survivors. Evidence-based scar management at present is challenging due to limitations of currently available validated tools to evaluate scar outcome and a lack of well-designed studies. Scar assessment is essential for diagnosis, monitoring, evaluation and therapeutic management of scar pathology. Currently most burn care facilities use subjective scar scales to monitor burn scar progression and to guide treatment. However, these scales are inherently prone to misclassification bias.

There is a lack of objective assessment in burn scar evaluation, suggesting a need for a reliable objective assessment tool. The DermaLab Combo® was identified as a multi-component device with the capacity to measure all four significant burn scar parameters. At present there is no accepted ‘gold standard’ method in burn scar assessment, and this introduces challenges for reliability and validity testing of new scar assessment scales and/or objective tools.

The overarching aim of the project was to translate an objective scar assessment method into a new tool for clinical burns practice. The specific objectives of this project were to identify an appropriate objective tool with capacity to measure all relevant burn scar parameters and to assess the reliability and validity of such a tool in burn scar assessment. The reliability of each of the DermaLab Combo® measurements of burn scar parameters (pigmentation, vascularity, pliability and thickness) was assessed and the interpretation and validity of these measurements were explored. The current method of scar assessment used at Royal Perth Hospital is a mVSS for which reliability had not been established. As a prerequisite for the reliability and validity
testing of the DermaLab Combo® measurements, it was necessary to establish the inter-rater reliability of this mVSS for use as the reference standard.

1.10 Study plan

This research project is presented in three phases:

Phase One

Chapter 1 – Literature review

To identify the potential objective device for translation to the clinical burns practice

Phase Two

Chapter 2 – Testing the inter rater reliability of the mVSS

To establish the inter-rater reliability of the mVSS method used at RPH and use as a reference standard to test the DermaLab Combo®

Chapter 3 – Testing the reliability of the DermaLab Combo®

To test the inter-rater and test-retest reliability of the DermaLab Combo® in measuring scar parameters analogous to the mVSS: pigmentation, vascularity, pliability and height

Chapter 4 – Exploring the interpretation of the DermaLab Combo® pigmentation and vascularity measurements in relation to the mVSS

To explore the interpretation and validity of the pigmentation (melanin) and vascularity (erythema) measurements of the DermaLab Combo®

Phase Three

Chapter 5 – Synthesis of project results, conclusions and recommendations

Overview of clinical studies, recommendations for future testing of the DermaLab Combo®, emerging themes, project strengths and limitations, conclusions and recommendations
Chapter 2
Chapter 2 – Inter-rater reliability of the Modified Vancouver Scar Scale (mVSS) used at Royal Perth Hospital

This chapter describes the clinical study conducted to establish the inter-rater reliability of the current scar assessment method used at RPH and is based on the following original article titled ‘A modified Vancouver Scar Scale linked with TBSA (mVSS-TBSA): Inter-rater reliability of an innovative burn scar assessment method’ published in Burns prior to the submission of this thesis (Gankande et al. 2013) [Appendix A].

2.1 Introduction

Cutaneous healing in all but the most superficial wounds results in a scar. Excessive or ‘hypertrophic’ scar formation is more common in post burn injuries than other traumatic injuries to the skin (Gangemi et al. 2010; Gangemi et al. 2008; Kaartinen et al. 2011) and one of the main objectives of treatment is to minimise scarring (Gangemi et al. 2010). All interventions from the point of injury aim to improve functionality, aesthetics and psychological outcomes to give a better quality of life after burn injury (Gangemi et al. 2008; Wiechman & Patterson 2008; Bloemen et al. 2011a).

Monitoring the progression of a scar using some form of scar assessment is essential to evaluate the impact of therapy and the need to modify treatment if necessary (Roques & Teot 2007). Scar assessments are also fundamental for measuring research outcomes, allowing the comparison of factors that may influence scar outcomes as well as the assessment of clinical interventions.

Scar assessment scales are subjective in nature and provide a qualitative measure of the scar by the person performing the assessment (clinician/observer/patient) and have low inter-rater reliability (Corica et al. 2006). Despite the limitations associated with scar scales, they are still used widely in the clinical setting because they are cheap, time-efficient and do not require much training (Brusselaers et al. 2010a).

The VSS is the first validated burn scar assessment scale and remains the most widely used scale within the clinical setting (Corica et al. 2006). In order to address issues associated with the use of the VSS, such as subjectivity and low inter-rater reliability, many authors have designed modifications to the VSS over the last few years. At present there are at least eight modified VSS in the published literature (Tyack et al. 2012).
Although the VSS and its subsequent modifications are a step towards providing a framework for outcome assessment, they only measure a single area within the scar and do not capture the variation in pathology across the surface area of the scar (Nedelec et al. 2008). We propose a different system of scar assessment based on the VSS the Baryza modification (mVSS) (Baryza & Baryza 1995). The mVSS-TBSA is a new method of application of the mVSS that links the mVSS with the assessment of % total body surface area (%TBSA) of the scar based on the ‘Rule of 9s’ (Smith et al. 2005). The proposed new application has been developed to provide both improved description of the ‘whole’ scar and more complete information to enable monitoring of scar progression over time. For example, if at one month post-injury a 10% TBSA scar contains an area of 2% TBSA with VSS score of 10, and at 6 months post-injury the area with VSS score of 10 has reduced to 0.5% TBSA, an improvement of 1.5% TBSA can be monitored using this mVSS-TBSA, even though the VSS score of the ‘worst’ scar area remains static at 10 (Figure 2-1).

Figure 2-1: Comparison of scores using conventional and new method mVSS-TBSA

A study to assess the inter-rater reliability of this new mVSS-TBSA was undertaken at the RPH Burn Outcome Centre, Western Australia (WA). The expectation was that three raters, with a range of duration of experience in scar assessment, could measure scar outcome reliably using the new method.
2.2 Method

2.2.1 Subjects

A group of 30 adult patients scheduled for scar assessments as a part of their routine follow-up care post burn injury were recruited. One index scar per patient (see Procedure) was assessed for a total of 30 scars. Subjects were all volunteers and met the following eligibility criteria: at least 18 years of age, presence of a healed immature burn scar less than 18 months post injury, scar area of at least 6 cm x 3 cm and able to provide informed written consent. Subjects who had other dermatological conditions in conjunction with their burn injury were excluded from the study. Written informed consent was obtained after providing subjects with a Patient Information Sheet and a verbal explanation. The study was approved by the Human Research Ethics Committees of RPH and UWA.

2.2.2 Study design

The study was designed as a prospective correlation study to assess the inter-rater reliability of the mVSS and %TBSA as part of the new mVSS-TBSA method. Three raters assessed 30 individual index scars on 30 patients, i.e., one scar per patient. Within each individual index scar the 'best' scar area of 3 cm x 3 cm and the 'worst' scar area of 3 cm x 3 cm were assessed using the mVSS. Therefore, the raters assessed 60 scar sites in total. Raters then allocated the total body surface area of the scar (%TBSA) to three mVSS categories (<5, 5-10, >10). Each rater was blinded to the results of other raters. The person performing the data analysis was blinded to the identity of each rater.

2.2.3 Raters

All raters had conducted at least 50 assessments using the new mVSS-TBSA prior to the commencement of the study. The raters included a senior occupational therapist with over five years of experience of the method, a second senior occupational therapist with over eight months experience of the method, and a clinician researcher with six months experience of the method. For the remainder of the paper, we will refer to these raters as R1, R2 and R3.
2.2.4 Procedure

All pressure garments and bandages were removed at least 15 minutes before the start of the first assessment. In the event that a patient had more than one scar the anatomical body part with the highest %TBSA scar (healed) was selected as the index scar for assessment. The index scar (right arm, left arm, back, chest, right leg, left leg, head, neck and groin) was identified by the most experienced Rater. A 3 cm x 3 cm scar area deemed to be the ‘best’ area of the scar (lowest mVSS score) and 3 cm x 3 cm area of the same scar deemed to be as the ‘worst’ area of the scar (highest mVSS score) were identified by each rater, and rated based on the mVSS (Baryza modification) (Baryza & Baryza 1995). This modification incorporates a plexiglass tool for blanching the scar and measuring scar height. Each rater then used the ‘Rule of 9s’ to estimate the scar area (%TBSA) and allocated the %TBSA to three mVSS severity categories (<5 mVSS, 5-10 mVSS, >10 mVSS). The three independent raters performed the scar assessments on each subject on the same day. The raters were blinded to the results recorded by other raters.

**Table 2-1: Numerical scale of mVSS used in the study**

<table>
<thead>
<tr>
<th>Pigmentation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = normal</td>
<td>1 = hypopigmentation</td>
</tr>
<tr>
<td>2 = mixed pigmentation</td>
<td>3 = hyperpigmentation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vascularity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = normal</td>
<td>1 = pink</td>
</tr>
<tr>
<td>2 = red</td>
<td>3 = purple</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pliability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = normal</td>
<td>1 = supple – flexible with minimal resistance</td>
</tr>
<tr>
<td>2 = yielding – giving way to pressure</td>
<td>3 = firm – inflexible, not easily moved, resistant to manual pressure</td>
</tr>
<tr>
<td>4 = banding – rope-like tissue that blanches with extension of scar</td>
<td>5 = contracture – permanent shortening of scar producing deformity or distortion.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Height</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = normal – flat</td>
<td>1 = 0 to 1 mm</td>
</tr>
<tr>
<td>2 = 1 to 2 mm</td>
<td>3 = 2 to 4 mm</td>
</tr>
<tr>
<td>4 = &gt; 4 mm</td>
<td></td>
</tr>
</tbody>
</table>
2.3 Data collection

Data was collected over a period of five weeks. The results were recorded in the Assessment Form which comprises the numerical mVSS scale (Table 2-1), a body map and a chart for allocation of the %TBSA (Appendix 1 – Supplementary Figure 1) (Smith et al. 2005). All forms were filed separately until the data collection was complete.

2.4 Data analysis

Descriptive statistics including age, gender, scar %TBSA, anatomical location of the scar, and time since injury, were calculated. Separate inter-rater reliability assessments were made for the ‘best’ and ‘worst’ scar areas for each mVSS component and the total mVSS score. The inter-rater reliability of determining the %TBSA of the whole scar and the allocation of the scar %TBSA to three severity categories (<5 mVSS, 5-10 mVSS, >10 mVSS) was also assessed. Data analyses were performed using SPSS version XX (Chicago, Inc.) and Stata V12 statistical software (StataCorp LP, College Station, Tex).

The Bland Altman method was used for continuous variables (total mVSS score, scar %TBSA and the allocation of scar %TBSA to the three severity categories) and plots of the difference in scores versus average of scores, for each pair of raters (R1 vs R2, R1 vs R3, and R2 vs R3) were generated with mean difference and 95% limits of agreement overlaid (Bland & Altman 1986). The 95% limits of agreement represent limits of how different the measurements of two observers could plausibly be for a subject and were calculated as the mean of the differences ±2 x standard deviation. The proportion of values falling within the 95% limits of agreement was reported for each rater combination. Plots were inspected for any tendency for the amount of variation to change with the magnitude of the measurement. If no such tendency was observed, inter-rater reliability was assessed using two-way random effects. ANOVA was used to derive the intra-class correlation coefficients (ICC) with 95% confidence intervals for each pair of raters.

The ICC is a measure of agreement that records the average similarity or reproducibility of raters’ actual scores on the ratings being compared (McDowell & Newell 1996). The ICC is a measure of the proportion of total variation that is between subjects variation. Therefore, a limitation of the ICC is that if variation due to observers is large compared with the variation between subjects then ICC will be closer to zero (Jones et al. 2011). Different guidelines exist for the interpretation of the ICC, however, one acceptable scale is that an ICC value of less than 0.40 indicates poor reproducibility or agreement, ICC values in the range of 0.40 to 0.75 indicate fair to good agreement, and an ICC value of greater than 0.75 demonstrates excellent agreement (Rosner 2006; Sampat et al. 2006).
For the nominal scales of the mVSS (pigmentation, vascularity, pliability, height), inter-rater reliability was assessed using the weighted kappa statistic ($k_w$) with quadratic weights (Streiner & Norman 1996) for each pair of raters (R1 vs R2, R2 vs R3 and R1 vs R3). The calculation of the kappa statistic is based on the difference between how much agreement is actually present (‘observed’ agreement) compared to how much agreement would be expected to be present by chance alone (‘expected’ agreement) (Streiner & Norman 1996). In addition to $k_w$, for height assessments in the ‘best’ scar area, observed and expected agreement, as well as the observed proportions of agreement for height category=0 and height category=1, as defined by Cicchetti and Feinstein (1990), were reported for rater combinations R1 vs R2 and R2 vs R3. The interpretation of $k_w$ values by Rosner (2006) is that 0-0.4 indicates marginal agreement, 0.4-0.75 is indicative of good agreement while a score over 0.75 demonstrates excellent agreement.

2.5 Results

2.5.1 Subjects

Among the 30 subjects, 22 (73.3%) were male and the median age of study subjects was 30.5 years (inter-quartile range [IQR], 26.7-53.0 years). The median scar %TBSA was 1.5% (IQR 1.0-4.8%) and the median time since injury among subjects was 3.0 months (IQR 1.8-6.5 months). The locations of the assessed scars were: arm (n=11), leg (n=12), chest (n=3), head and neck (n=2), and back (n=2).

2.5.2 mVSS scores

Figure 2-2 summarises the total mVSS scores derived by all raters in the ‘best’ and ‘worst’ scar areas. There were only a few scores above mVSS 10, with the majority of mVSS scores being in the <5 and 5-10 mVSS categories. As expected, there were more scores in the 5-10 and >10 mVSS categories in the ‘worst’ scar area compared to the ‘best’ scar area.
2.5.3 Inter-rater reliability

Table 2-2 presents pair-wise weighted kappa statistics for each mVSS component in the ‘best’ and ‘worst’ area of the scar. Rater agreement with respect to the mVSS components was variable. Pigmentation assessment was the least reliable with ‘marginal’ agreement demonstrated for the ‘worst’ scar area ($k_w = 0.30, 0.02, 0.33$ for each rater pair) and ‘good’ agreement ($k_w = 0.57, 0.45, 0.54$ for each rater pair) demonstrated in the ‘best’ scar area. Pliability assessment achieved ‘excellent’ agreement with $k_w$ scores of $0.83, 0.84$ and $0.82$ for the ‘best’ scar area and $k_w$ scores of $0.77, 0.86$ and $0.84$ in the ‘worst’ scar area. Height assessment demonstrated ‘good to excellent’ agreement in the ‘worst’ scar area ($k_w = 0.85, 0.72$ and $0.74$). In the ‘best’ scar area, all height measurements recorded were either category 0 or 1, with R2 recording all height measurements as category 0. For rater combinations R1 vs R2 and R2 vs R3, 25/30 pair-wise height measurements recorded scores of zero, with observed and expected agreement of $83.3\%$. The $k_w$ scores ($0.00$) would suggest poor agreement for these rater combinations when accounting for chance, however the observed proportions of agreement for height category=0 and category=1 were $0.91$ and $0$ respectively (Altman 1991; Feinstein & Cicchetti 1990). For rater pair (R1 vs R3) the $k_w$ score demonstrated ‘excellent’ agreement ($k_w = 0.76$). Vascularity assessment achieved ‘good to excellent’ agreement in the 'worst' scar area, with $k_w$ scores of $0.64-0.76$, and $k_w$ scores of $0.44-0.71$ in the ‘best’ scar area, indicative of good agreement.
Table 2-2: Weighted Kappa ($k_w$) values for individual rater pairs in 'best' and 'worst' scar areas

<table>
<thead>
<tr>
<th>Best area of the scar</th>
<th>Worst area of the scar</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=30</td>
<td>n=30</td>
</tr>
<tr>
<td>$k_w$</td>
<td>$k_w$</td>
</tr>
<tr>
<td>R1 vs R2</td>
<td>R2 vs R3</td>
</tr>
<tr>
<td>Pigmentation</td>
<td>0.57</td>
</tr>
<tr>
<td>Vascularity</td>
<td>0.44</td>
</tr>
<tr>
<td>Pliability</td>
<td>0.83</td>
</tr>
<tr>
<td>Height</td>
<td>nd*</td>
</tr>
</tbody>
</table>

*nd*=indeterminate as one observer showed no variation

Bland Altman plots for the continuous variables (total mVSS score in ‘best’ and ‘worst’ areas, scar %TBSA and allocation of scar %TBSA in three severity categories: <5 mVSS, 5-10 mVSS, >10 mVSS) were inspected for systematic patterns. No tendency for variation to change with the magnitude of the measurement was observed (Appendix 2). For all Bland Altman plots of difference in scores versus the mean of the score for each pair, almost all values fell within the limits of agreement.

ICCs and Bland Altman values (mean difference, upper and lower 95% limits of agreement, proportion within limits of agreement) for pairs of raters for total mVSS scores and %TBSA and mVSS category allocation are presented in Table 2-3 and Table 2-4, respectively. The total mVSS score had ‘good’ rater agreement in the ‘best’ scar area (ICC 0.65-0.73) for all the rater combinations. In the ‘worst’ area of the scar, there was excellent rater agreement (ICC 0.85-0.88) for all the rater combinations. The rater agreement in determining the %TBSA of the index scar was ‘excellent’ (ICC 0.90) for all rater combinations. The agreement between pairs of raters for allocation of the scar %TBSA to mVSS categories (<5 mVSS, 5-10 mVSS, >10 mVSS) varied (Table 2-4). Results of the ICC supported good agreement for R1 vs. R3 (ICC 0.63) in allocating %TBSA to the <5 mVSS category and ICC results demonstrated ‘excellent’ agreement for R1 vs R2 and R2 vs R3 (ICC 0.80, 0.78). The allocation of scar %TBSA to the 5-10 mVSS category showed ‘fair to good’ agreement using ICCs (ICC 0.42–0.74).

In Table 2-3, the Bland Altman mean differences for total mVSS scores varied from -0.50 to 0.57 and -1.0 to 0.37 for ‘best’ and ‘worst’ scar area, respectively. The rater pair R1 vs R2 showed the greatest limits of agreement for total mVSS scores for both the ‘best’ and ‘worst’ scar area of -2.15 to 3.28 and -2.81 to 3.54, respectively. In Table 2-4, for the scar %TBSA estimation, the mean differences varied from -0.51 to 0.20. Rater pair R2 vs R3 showed the greatest limits of agreement for scar %TBSA estimation of -3.99 to 2.97. With respect to
allocation of the scar %TBSA to category <5 mVSS and 5-10 mVSS, the mean differences varied from -0.32 to 0.09 and -0.33 to 0.29, respectively. Rater pair R2 vs R3 showed the greatest limits of agreement for allocation of scar %TBSA to <5 mVSS category (-3.82 to 3.19). Rater pair R1 vs R2 had the greatest limits of agreement for allocation of scar %TBSA to 5-10 mVSS (-1.96 to 2.54).

Table 2-3: ICC and Bland Altman values (mean score difference, 95% limits of agreement) for total mVSS for “best” and “worst” area of scar for rater pairs

<table>
<thead>
<tr>
<th></th>
<th>R1 vs R2</th>
<th>R2 vs R3</th>
<th>R1 vs R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total mVSS Score (Best scar area)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICC (95% CI)</td>
<td>0.65 (0.38, 0.81)</td>
<td>0.72 (0.50, 0.86)</td>
<td>0.73 (0.51, 0.86)</td>
</tr>
<tr>
<td>Mean Difference (95% CI)</td>
<td>0.57 (0.06, 1.07)</td>
<td>-0.50 (-0.90, -0.10)</td>
<td>0.07 (-0.41, 0.55)</td>
</tr>
<tr>
<td>Lower limit of agreement (95% CI)</td>
<td>-2.15 (-3.02, -1.27)</td>
<td>-2.65 (-3.34, -1.95)</td>
<td>-2.50 (-3.33, -1.67)</td>
</tr>
<tr>
<td>Upper limit of agreement (95% CI)</td>
<td>3.28 (2.40, 4.16)</td>
<td>1.65 (0.95, 2.34)</td>
<td>2.64 (1.81, 3.47)</td>
</tr>
<tr>
<td>Proportion within limits of agreement</td>
<td>28/30 (93%)</td>
<td>29/30 (97%)</td>
<td>29/30 (97%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>R1 vs R2</th>
<th>R2 vs R3</th>
<th>R1 vs R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total mVSS Score (Worst scar area)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICC (95% CI)</td>
<td>0.85 (0.71, 0.93)</td>
<td>0.86 (0.73,0.93)</td>
<td>0.88 (0.76, 0.94)</td>
</tr>
<tr>
<td>Mean Difference (95% CI)</td>
<td>0.37 (-0.23, 0.96)</td>
<td>-1.00 (-1.56, -0.44)</td>
<td>-0.63 (-1.17, -0.09)</td>
</tr>
<tr>
<td>Lower limit of agreement (95% CI)</td>
<td>-2.81 (-3.83,- 1.78)</td>
<td>-4.02 (-4.99, -3.04)</td>
<td>-3.53 (-4.47, -2.68)</td>
</tr>
<tr>
<td>Upper limit of agreement (95% CI)</td>
<td>3.54 (2.51, 4.57)</td>
<td>2.02 (1.04, 2.99)</td>
<td>2.27 (1.33, 3.20)</td>
</tr>
<tr>
<td>Proportion within limits of agreement</td>
<td>27/30 (90%)</td>
<td>29/30 (97%)</td>
<td>29/30 (97%)</td>
</tr>
</tbody>
</table>
Table 2-4: ICC and Bland Altman values (mean score difference, 95% limits of agreement) for scar %TBSA and allocation of scar %TBSA to <5 mVSS and 5-10 mVSS categories for rater pairs

<table>
<thead>
<tr>
<th>% TBSA</th>
<th>R1 vs R2</th>
<th>R2 vs R3</th>
<th>R1 vs R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC (95% CI)</td>
<td>0.95 (0.90, 0.97)</td>
<td>0.91 (0.82, 0.95)</td>
<td>0.96 (0.93, 0.98)</td>
</tr>
<tr>
<td>Mean Difference (95% CI)</td>
<td>0.20 (-0.26, 0.65)</td>
<td>-0.51 (-1.16, 0.14)</td>
<td>-0.31 (-0.73, 0.10)</td>
</tr>
<tr>
<td>Lower limit of agreement (95% CI)</td>
<td>-2.26 (-3.05, -1.46)</td>
<td>-3.99 (-5.11, -2.86)</td>
<td>-2.53 (-3.23, -1.80)</td>
</tr>
<tr>
<td>Upper limit of agreement (95% CI)</td>
<td>2.65 (1.86, 3.44)</td>
<td>2.97 (1.84, 4.09)</td>
<td>1.89 (1.18, 2.60)</td>
</tr>
<tr>
<td>Proportion within limits of agreement</td>
<td>29/30 (97%)</td>
<td>29/30 (97%)</td>
<td>28/30 (93%)</td>
</tr>
</tbody>
</table>

<5 mVSS
| ICC (95% CI)         | 0.80 (0.62-0.89)      | 0.78 (0.59-0.89)      | 0.63 (0.35-0.80)      |
| Mean Difference (95% CI) | 0.09 (-0.49-0.67)   | -0.32 (-1.01-0.38)   | -0.23 (-0.90-0.43)   |
| Lower limit of agreement (95% CI) | -2.82 (-3.82, -1.82) | -3.82 (-5.01, -2.62) | -3.51 (-4.66, -2.36) |
| Upper limit of agreement (95% CI) | 3.00 (2.01, 4.00)   | 3.19 (1.99, 4.39)    | 3.05 (1.89, 4.19)    |
| Proportion within limits of agreement | 25/27 (93%)       | 25/27 (93%)          | 25/26 (96%)          |

5-10 mVSS
| ICC (95% CI)         | 0.74 (0.52-0.87)      | 0.42 (0.08-0.66)      | 0.54 (0.23-0.75)      |
| Mean Difference (95% CI) | 0.29 (-0.39, 0.97)   | -0.33 (-0.68, 0.01)  | -0.09 (-0.70-0.53)   |
| Lower limit of agreement (95% CI) | -1.96 (-3.13, -0.78) | -1.35 (-1.95, -0.76) | -2.11 (-3.18, -1.05) |
| Upper limit of agreement (95% CI) | 2.54 (1.36, 3.72)   | 0.69 (0.19, 1.28)    | 1.95 (0.89, 3.02)    |
| Proportion within limits of agreement | 12/13 (92%)       | 11/11 (100%)         | 12/13 (92%)         |

Table 2-5: Clinically significant misclassification rates of individual rater pairs for ‘best’ and ‘worst’ areas of the scar

<table>
<thead>
<tr>
<th>Pair wise misclassification rates†</th>
<th>‘Best’ Area of Scar</th>
<th>‘Worst’ Area of Scar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vascularity*</td>
<td>4/90</td>
<td>3/90</td>
</tr>
<tr>
<td>Pliability*</td>
<td>2/90</td>
<td>8/90</td>
</tr>
<tr>
<td>Height*</td>
<td>0/90</td>
<td>0/90</td>
</tr>
<tr>
<td>Total mVSS score**</td>
<td>4/90</td>
<td>12/90</td>
</tr>
<tr>
<td>For the whole scar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%TBSA estimation***</td>
<td>6/90</td>
<td></td>
</tr>
<tr>
<td>&lt;5 mVSS***</td>
<td>17/90</td>
<td></td>
</tr>
<tr>
<td>5-10 mVSS***</td>
<td>7/90</td>
<td></td>
</tr>
</tbody>
</table>

† Total pair-wise Rater tests = 90 (three assessments per patient for 30 patients)
*Difference of two units in the sub-scale score
**Misclassification on basis of <5 and ≥ 5 mVSS
*** Misclassification on basis of >1 %TBSA
2.6 Discussion

This study assessed the inter-rater reliability of the elements of the new mVSS-TBSA method, namely, mVSS scoring and %TBSA allocation in three categories of scar severity (<5 mVSS, 5-10 mVSS, >10 mVSS). For the ‘worst’ scar area, there was ‘good to excellent’ inter-rater reliability for three of the individual mVSS components (vascularity, pliability and height), with ‘marginal’ reliability demonstrated for pigmentation assessment. ‘Good to excellent’ inter-rater reliability was shown for all individual mVSS components (pigmentation, vascularity, pliability and height) in ‘best’ area of the scar.

Assessment of total mVSS scores showed ‘fair to good’ inter-rater reliability in ‘best’ scar area while in the ‘worst’ area there was ‘good to excellent’ inter-rater reliability for the same. Overall, higher levels of inter-rater agreement were observed for total mVSS measurements of the ‘worst’ area of the scar than the ‘best’ area. With respect to the estimation of scar %TBSA, there was ‘excellent’ inter-rater agreement. The allocation of scar %TBSA to mVSS severity categories demonstrated varying levels of reliability. The allocation of scores to the category <5 mVSS showed ‘good to excellent’ reliability. Allocation of scar %TBSA to the 5-10 mVSS category demonstrated ‘fair to good’ reliability (ICC 0.42-0.74), however only 11 to 13 scars tested (Table 2-4) fell within this category and this small sample size may have impacted the ICC scores generated.

The method proposed in this study links the mVSS (Baryza modification) to the %TBSA of the assessed scar. This enables scars to be monitored in terms of change in scar %TBSA in three mVSS categories. Using the mVSS alone, a selected area of the scar is assessed, and the maximum mVSS score may not change over the monitoring period. Results of this study demonstrate that raters can reliably determine and allocate scar %TBSA to two broad categories of scar severity (<5 mVSS, ≥5 mVSS). In clinical practice it is often the situation that the personnel involved in burn scar assessment have varying levels of experience. The results of this study support that someone with a minimum of six months training and experience in scar assessment can reliably conduct scar assessments using the mVSS-TBSA method.

Inter-rater reliability of pigmentation assessment was ‘marginal’. The assessment of pigmentation is highly subjective, whereby raters need to make judgements on the 'brownness' of the scar. In addition, pigmentation is confounded by the colour of the scar due to vascularity even when ‘blanched’, since the degree of blanching depends on the pressure on the plexiglass by each Rater. Forbes-Duchart and others (2007) also found that pigmentation was not a reliable parameter.

Height assessment in the 'best' area of the scar for two of the rater combinations reported high observed agreement but low $k_w$ scores ($k_w$ 0.00), a paradox that has been described in the
literature (Feinstein & Cicchetti 1990; Altman 1991; Viera & Garrett 2005). The paradox observed for ‘best’ scar area height measurement reflects the low prevalence of rating 1 and high prevalence of rating 0 (normal skin), with a high observed proportion of agreement (Feinstein & Cicchetti 1990; Viera & Garrett 2005). Rater pair (R1 vs R3) showed ‘excellent’ rater agreement for height assessment in the ‘best’ scar area. In the ‘worst’ scar areas, all rater combinations demonstrated ‘good to excellent’ rater agreement.

Draaijers and others (2004c) have shown that with three raters the ICC for total mVSS score using their modification of the VSS was ‘excellent’ (ICC 0.86-0.88 at 95% CI). Our findings for the inter-rater reliability for assessment of the ‘worst’ area of the scar (ICC 0.85-0.88) are similar to those reported by Draaijers and others (2004c). It is reasonable to assume that the area of scar assessed in Draaijer and others (2004c) was the ‘worst’ scar area. The study presented here provides important new data on how the inter-rater reliability of mVSS components depends on the severity of the scar area being assessed.

A limitation of this study was that there was insufficient data to complete a meaningful analysis on mVSS scores over 10. Only three of the study sample of 30 subjects had mVSS scores in the category mVSS >10 according to one or more of the raters. Further assessments that incorporate greater numbers of subjects with areas of poor scar outcomes are required to establish the reliability of scar %TBSA allocation to >10 mVSS. The sample size chosen for this study was based on recommendations by Walter and others (1998). However, a larger sample size would have enabled determination of reliability estimations with greater statistical confidence (Streiner & Norman 1996). Whilst the patient sample used in this study reflects the patient population of adults at our burns unit, a larger sample size may have provided more patients with diverse scar outcomes. The ICC values derived in this study suggest that the inter-rater reliability of the mVSS-TBSA method used is good. However, the results of the Bland Altman analyses revealed distributions of differences between rater scores that would question the level of reliability derived by the ICC analyses. Interpretation of the mean and limits of agreement of the difference in rater scores derived by the Bland Altman method relies on the clinical context, and acceptable clinical agreement cannot be based purely on statistics (Walter et al. 1998). In the clinical context of scar assessment, the differences in scores arising from different raters are large enough to misclassify scar outcome in some cases.

The authors acknowledge that many of the components of the mVSS are ordered categorical variables and the validity of combining such scores to a total mVSS score is questionable. It is also debatable whether the total mVSS score can be regarded as a continuous measurement, as has been done in previous published papers. However, peer-reviewed publications of inter-rater-reliability of the VSS and its modifications are based on total VSS score to determine scar
outcome. To be consistent with previous use and publications of the VSS we have reported mVSS total scores.

The original VSS, developed by Sullivan in 1990 (Sullivan et al. 1990) was designed to rate scars approximately 4 cm x 4 cm in size and one year post burn capturing four scar parameters: pigmentation, vascularity, pliability and height (Sullivan et al. 1990; Roques & Teot 2007; Nedelec et al. 2008). Despite these specifications it is broadly utilised to score scars of varying level of size and maturity. Furthermore, the VSS in its original version, and the many modifications since then, fail to capture the clinically relevant variation in scar outcome within each scar. The mVSS-TBSA method described in this study was specifically developed to take this variation in scar outcome into consideration, based on scales that are already familiar to burns clinicians and researchers. Adding a spatial dimension to scar assessment enhances the currently available information from scar assessments for clinicians beyond an indication of the severity of the ‘worst’ scar area. The mVSS-TBSA method could be used with a variety of reliable scar assessment tools that are designed to monitor specific areas of scarring to provide a better overall perspective of scarring.

2.7 Conclusion

This innovation linking the scar %TBSA to the mVSS provides a new method for identifying changes in scar outcome for improved clinical surveillance and decision-making, as well as for research purposes. From these data we conclude that there is evidence to support the reliability of this new mVSS-TBSA method in scar assessment. However, given the wide limits of agreement found in this study there is potential to misclassify scar outcome in some cases. Efforts to improve the inter-rater reliability of pigmentation and vascularity components using more objective tests will further improve the inter-rater reliability of total mVSS score and have a positive impact on the utility of the mVSS-TBSA.
Chapter 3
Chapter 3 – Reliability of scar assessments performed with an integrated skin testing device – the DermaLab Combo®

This chapter describes the clinical study conducted to establish the reliability of the DermaLab Combo® device for measuring scar parameters analogous to the mVSS scar assessment method used at RPH and is based on the following original article titled ‘Reliability of scar assessments performed with an integrated skin testing device – the DermaLab Combo®’ published in Burns prior to the submission of this thesis (Gankande et al. 2014) [Appendix B].

3.1 Introduction

Scarring is a consequence of burn injury with the potential for long-term physical and psychological impacts on patients (Van Loey & Von Son 2003). Systematic scar assessment enables clinicians to monitor scar progression and researchers to evaluate interventions and factors influencing scar outcome.

Current scar assessment methods used in clinical practice are subjective and may have low inter-rater reliability (Fearmonti et al. 2010). The inter-rater reliability of the mVSS, for example, varies depending on which scar component is being measured (Gankande et al. 2013). Objective scar measurements are likely to provide more reliable data for clinical and research purposes, enabling more robust comparisons within and between patients (Roques & Teot 2007; Durani et al. 2009; Kaartinen et al. 2011).

Scar assessment includes measurement of four key components – pigmentation, vascularity, pliability and height (thickness) of the scar. Many individual devices have been tested to obtain objective measurements of these components, but none have been broadly adopted within a clinical setting. The most likely reason is that most devices only measure one individual component (Idriss & Maibach 2009; Brusselaers et al. 2010b; Oliveira et al. 2005). For example, in Oliveira and others (2005), objective measurements for the four scar components were obtained using four separate devices.

Scar measuring devices can be broadly categorised into three groups, assessing colour (pigmentation and vascularity), pliability (elasticity) and height (thickness) (Brusselaers et al. 2010b). To date no device capable of measuring all four scar components has been tested for reliability (Brusselaers et al. 2010b; Idriss & Maibach 2009). The DermaLab Combo® (Cortex Technologies, Denmark) (Cortex Technologies 2011) is a commercially-available skin testing device that has the potential capability to measure all four scar assessment components. It was
primarily designed for skin testing in the cosmetic field and does not claim to be a medical device. However, it is a high-specification device based on proven technologies (see Methods).

This study was designed to test the reliability (inter-rater reliability and test-retest reliability) of the DermaLab Combo® device to measure the four components of the mVSS in burn scars: pigmentation, vascularity, pliability and height (thickness) (Baryza & Baryza 1995). The study was conducted at RPH burns outpatient clinic, WA.

3.2 Methods

3.2.1 Subjects

Thirty adult patients who attended the RPH burns outpatient clinic for routine follow-up assessments and were able to attend a follow-up appointment within 10 days of the first assessment were recruited consecutively. Participation was voluntary. Subjects were included in the study if they met the following inclusion criteria: at least 18 years of age; able to provide informed written consent; had a healed burn scar with an area of at least 3 cm x 6 cm; the scar was a minimum of three months post burn injury; and there was a contralateral, normal skin area of at least 3 cm x 3 cm with the same degree of sun exposure.

Subjects were excluded from the study if they could not provide informed written consent. If the subject wore a pressure garment, the rater responsible for recruiting the subject made a subjective assessment on the level of sun exposure of the contralateral normal skin and asked supplementary questions to confirm a similar level of sun exposure on both scar and contralateral normal skin areas.

All subjects were provided with a patient information sheet and a verbal explanation of the study after which written informed consent was sought.

3.2.2 Study design

The study was a single-arm observational study using three independent raters. For the inter-rater reliability study, all three raters made measurements of the scar and contralateral normal skin (control) areas for each parameter on the same day. For the test-retest reliability study, the same three raters made the same measurements on a second occasion within 10 days of the initial assessment. The study was approved by RPH and UWA Human Research Ethics Committees.
3.2.3 Raters

The three raters had comprehensive training in the use of the DermaLab Combo® under the supervision of an experienced user of the DermaLab Combo® (HW). Each training session was three hours in duration and the raters undertook three sessions of training over a three week period. The training session consisted the raters familiarising themselves with the use of the probes and the study protocol to measure the four components of scar assessment on each other. Subsequent to this training, each rater used the DermaLab Combo® device on patients with burn scars to conduct at least 10 independent assessments of scars using the study protocol prior to the start of the study. In this study the raters are referred to as raters R1, R2 and R3.

3.2.4 Materials

The DermaLab Combo® was used to assess melanin (pigmentation), erythema (vascularity), elasticity (pliability) and thickness (height) of post-burn scars. The DermaLab Combo® consists of a main unit with screen and multiple separate probes. The three probes used in this study had separate channel connectors and were each attached to the main unit with separate tubes (Figure 3-1).

3.2.4.1 The colour probe

The colour measurement of the DermaLab Combo® is based on the principle of narrow-band reflectance spectrophotometry (550 nm ±30 nm and 660 nm ±60 nm for haemoglobin [erythema or vascularity] and melanin [pigmentation] respectively). The colour probe has an optical focussing on 7 mm diameter target area with a clear front for accurate positioning and is illuminated by two angled white light emitting diode (LED) lights. The probe displays four readings (three individual measurements with their average) separately for erythema (vascularity) and melanin (pigmentation) using Commission Internationale de l'Eclairage (CIE) – luminance (L), red-green axis (a) and blue-yellow axis (b) [CIELab] values.

3.2.4.2 The elasticity probe

The elasticity measurement of the DermaLab Combo® is based on the principle of vertical suction applied on the surface of scar. The probe has a measuring aperture of 10 mm diameter and adheres to the skin by a double adhesive sticker. Elasticity measurements are expressed in terms of Young’s modulus in megaPascals (MPa). For the purpose of this paper, elasticity measurements will be referred to as ‘pliability’.
3.2.4.3 The skin thickness probe

The DermaLab Combo® skin thickness measurement probe is based on the principle of high frequency ultrasound (20 MHz) and has a resolution of 60 x 200 micrometres with a penetration capacity of 3.4 mm with a fully adjustable gain settling ±10dB). The probe has a rotating transducer, scan length 17 mm, footprint 11 mm, and the capacity to display actual and stored measurements side-by-side.

In contrast to the VSS height measurement that measures scar height above the pre-injury ‘baseline’ surface of the skin, this probe measures the thickness of the whole scar including the scar above and below the baseline surface of the skin. For the purpose of this paper, scar height measurements will be referred to as ‘thickness’.

Figure 3-1: The DermaLab Combo® device and example outputs for the four parameters – melanin, erythema, thickness and elasticity (adapted from www.cortextechnologies.dk)

3.3 Procedure

3.3.1 Inter-rater reliability

All pressure garments and bandages were removed at least 15 minutes before the start of the assessment to reduce any blanching effect that may influence measurements. All assessments were carried out in the same room where the subjects were exposed to similar environmental conditions at the time of assessment. All assessments were carried out with the subjects in a sitting position. The anatomical body part containing the scar was placed in a nondependent position; if the scar was on a leg – both legs were elevated to rest on another chair, if the scar was on an arm – both arms were rested parallel to each other on a table.
One index scar of at least 6 cm x 3 cm in size was identified per eligible participant. The index scar was defined as the scar with the largest percentage of body surface area on a body site (head/neck, chest, back, left arm, right arm, left leg and right leg) with a contralateral anatomically-matched normal skin area available (minimum surface area 3 cm x 3 cm). Within the index scar, one nominated rater selected a 3 cm x 3 cm scar area deemed to be the ‘best’ area of the scar (lowest mVSS score) and a 3 cm x 3 cm area of the same scar deemed to be the ‘worst’ area of the scar (highest mVSS score) using the mVSS (Baryza & Baryza 1995). These two areas were marked by this rater with a semi-permanent skin marker and photographed. A 3 cm x 3 cm contralateral anatomically-matched normal skin area was also identified and marked.

All three raters assessed these three marked areas using the DermaLab Combo®. Each scar component was assessed at three sites within the identified 3 cm x 3 cm scar areas (see Figure 3-2). These individual measurements as well as the average of the three measurements were recorded and used in the analysis. The three independent raters performed the scar assessments on each subject on the same day. In the case of pliability measurement, suction applied by the DermaLab Combo® elasticity probe alters the biomechanical properties of the skin, hence repeat measurements on the same location are not recommended by the manufacturer at less than one hour intervals. Due to this extended wash-out period the elasticity probe was used only by one rater (R1) and was excluded from the inter-rater reliability analysis.

![Diagram of scar measurement protocol](image)

A = ‘Best’ area of scar  
B = ‘Worst’ area of scar

The circles numbered 1, 2 and 3 represent the sites for each individual measurement (probe placement). Each rater estimated the locations based on the marked square. The locations themselves were not marked, as this would interfere with the readings.

**Figure 3-2: Sampling protocol within marked 3 cm x 3 cm square used for measurement**
### 3.3.2 Test-retest reliability

All study subjects were asked to attend a repeat assessment within 10 days of the initial baseline assessment. Two or three of the original raters performed the repeat DermaLab Combo® scar assessment on each subject. The repeat assessment was carried out under the same conditions as the initial assessment. Subjects were asked about anything that may have affected their scar between the time of the initial baseline assessment and the repeat assessment (for example, use of different skin moisturiser, a new therapeutic measure such as a different pressure garment or excessive sun exposure). If the three marked scar areas were not visible they were re-marked with the semi-permanent skin marker with the help of the baseline photographs.

As for the initial assessment, each scar parameter was assessed on three sites within each marked area (Figure 3-2) and the individual measurements as well as the average of the three measurements were recorded. The raters performed the repeat scar assessments on each participant on the same day. Due to the extended wash-out period (one hour) with the elasticity probe, only R1 conducted the repeat pliability assessment.

### 3.4 Data collection and analysis

Data was collected over a six month period, from July 2012 to January 2013. Each rater was blinded to the results of the other raters and independently recorded results in a separate Scar Assessment Form (Appendix 3). Each rater recorded four measurements on the data collection sheet for each parameter in the three 3 cm x 3 cm areas: one reading for each measurement site and the average of the three readings. All forms were filed separately until the data collection was complete.

Throughout the paper we refer to the 3 cm x 3 cm squares marking the ‘best’ and ‘worst’ parts of the index scar and contralateral normal skin as ‘areas’ and the probe placement locations for measurements within the areas as ‘sites’.

For the purpose of the reliability analyses the readings for each site were considered as independent units for analysis. Analyses were also conducted using the average measurement of the three sites in the three 3 cm x 3 cm areas for inter-rater reliability. All data analyses were performed using SPSS version XX (Chicago, Inc.) and Stata V12 statistical software (StataCorp LP, College Station, Tex).
3.4.1 Inter-rater reliability

Inter-rater reliability is the extent of agreement among raters scoring the same subjects under the same conditions (Gwet 2008). The inter-rater reliability of individual measurements and the average measurement for each component were calculated for each rater pair (R1 vs R2, R1 vs R3 and R2 vs R3) for the ‘best’ and ‘worst’ area of each index scar and the contralateral normal skin area on each subject.

Measurements obtained with the DermaLab Combo® device were continuous variables. The Bland Altman (BA) method and plots of the difference in measurements versus average of scores, for each rater pair were generated with mean difference and 95% limits of agreement. BA plots and results were examined for any patterns of variation to change with the magnitude of the measurement. In the absence of such patterns inter-rater reliability was established using two-way random effects ANOVA to derive the ICC with 95% confidence intervals (CI) for all rater combinations. These methods are fully described in Gankande and others (2013). ICC were interpreted using the Rosner interpretation (0-0.40: ‘marginal’ agreement, >0.40-0.75: ‘good’ agreement and >0.75: ‘excellent’ agreement) (Rosner 2006).

3.4.2 Test-retest reliability

Test-retest reliability refers to the reliability or the consistency of a measurement over time (Kline 1986). Repeat scar assessments were done within 10 days of the initial baseline assessment, during which time no significant change in scar was anticipated. ICC was used as the measure of test-retest reliability (Bland & Altman 1986; Yen & Lo 2002).

Test-retest reliability of individual site measurements for each component were calculated for the ‘best’ and ‘worst’ area of each index scar and the contralateral normal skin area on each subject.

As for inter-rater reliability, the BA method was applied and test-retest reliability was assessed using two-way random effects ANOVA to derive ICC with 95% confidence intervals. ICC was interpreted using the Rosner interpretation (Rosner 2006).
3.5 Results

3.5.1 Subjects and descriptive statistics

Of the 30 subjects, 12 (40%) were female and 18 (60%) were male. The median age of study subjects was 43 years (IQR: 25.0-54.3 years, minimum (min)-maximum (max): 18-81 years). The median time since injury among subjects was 4.5 months (IQR: 3-6 months and min-max: 3-648 months). The locations of the assessed scars were: arm (n=12), leg (n=13), chest (n=3), abdomen (n=1) and back (n=1). Twenty-seven out of the 30 subjects had a mVSS assessment performed. The median mVSS score in the ‘best’ area of scar was 2 (IQR: 1-2 and min-max: 0-7) and 3 (IQR: 2-5; min-max: 1-10) in the ‘worst’ area of scar. The distributions of the scar component measurements at baseline for pigmentation, vascularity, thickness and pliability are described in Table 3-1.

Table 3-1: Range of scar measurements at baseline for pigmentation, vascularity, thickness and pliability measured with the DermaLab Combo® for the 30 study subjects

<table>
<thead>
<tr>
<th>Component</th>
<th>Minimum measurement</th>
<th>Maximum measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>‘Best’ site of scar</td>
<td></td>
</tr>
<tr>
<td>Pigmentation</td>
<td>26.00</td>
<td>93.00</td>
</tr>
<tr>
<td>Vascularity</td>
<td>10.00</td>
<td>45.00</td>
</tr>
<tr>
<td>Thickness</td>
<td>1001.00</td>
<td>2109.00</td>
</tr>
<tr>
<td>Pliability</td>
<td>177.00</td>
<td>3240.00</td>
</tr>
<tr>
<td></td>
<td>‘Worst’ site of scar</td>
<td></td>
</tr>
<tr>
<td>Pigmentation</td>
<td>28.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Vascularity</td>
<td>7.00</td>
<td>51.00</td>
</tr>
<tr>
<td>Thickness</td>
<td>1132.00</td>
<td>2433.00</td>
</tr>
<tr>
<td>Pliability</td>
<td>264.00</td>
<td>9438.00</td>
</tr>
<tr>
<td></td>
<td>Normal skin site</td>
<td></td>
</tr>
<tr>
<td>Pigmentation</td>
<td>26.00</td>
<td>66.00</td>
</tr>
<tr>
<td>Vascularity</td>
<td>9.00</td>
<td>43.00</td>
</tr>
<tr>
<td>Thickness</td>
<td>342.00</td>
<td>1870.00</td>
</tr>
<tr>
<td>Pliability</td>
<td>246.00</td>
<td>4704.00</td>
</tr>
</tbody>
</table>
3.5.2 Inter-rater reliability

ICC and 95% CI for all rater pairs (R1 vs R2, R1 vs R3 and R2 vs R3) for pigmentation, vascularity and thickness measurements are presented for individual measurements (Table 3-2) and average measurements (Table 3-3).

Pigmentation had ‘excellent’ inter-rater reliability in all the tested areas for all rater pairs, with ICC of 0.98, 0.94 and 0.95 in the ‘best’ area of the scar, 0.96, 0.95 and 0.97 in the ‘worst’ area of the scar and 0.95, 0.92 and 0.91 in the contralateral normal skin area. When the average measurements for pigmentation were used in the analysis the inter-rater reliability was marginally greater among all rater combinations.

Vascularity showed ‘good’ to ‘excellent’ inter-rater reliability in all three areas measured. ICC of 0.74, 0.66 and 0.78 in the ‘best’ scar area and 0.84, 0.67 and 0.73 in the ‘worst’ scar area correspond to ‘good’ to ‘excellent’ inter-rater reliability. In the contralateral normal skin area the inter-rater reliability was only ‘good’ and the ICC varied between rater pairs. Rater pairs R1 vs R2 and R1 vs R3 had higher inter-rater reliability (ICC 0.73) than the rater pair R2 vs R3 (ICC 0.54). As with pigmentation, there was marginally higher inter-rater reliability for vascularity when the average measurements were used in the analysis.

For about one third of the 30 subjects (11/30), the thickness measurement in the ‘worst’ area of scar reached the maximum thickness of 2.5 mm. Only one of the 30 subjects reached the maximum thickness in the ‘best’ area of the scar. When thickness was able to be measured, it achieved ‘excellent’ inter-rater reliability in all the tested areas for all rater pairs with ICC values of 0.95, 0.86 and 0.93 in the ‘best’ area of scar, 0.95, 0.91 and 0.96 in the ‘worst’ area of scar and 0.94, 0.92 and 0.95 in the contralateral normal skin area. Again, there was marginally higher inter-rater reliability when the average measurements were used in the analysis.
<table>
<thead>
<tr>
<th>Component</th>
<th>R1 vs R2‡</th>
<th>R1 vs R3‡</th>
<th>R2 vs R3‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC (95%CI)†</td>
<td>ICC (95%CI)†</td>
<td>ICC (95%CI)†</td>
</tr>
<tr>
<td><strong>‘Best’ scar site</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation (n=90)</td>
<td>0.98 (0.96, 0.98)</td>
<td>0.94 (0.90, 0.96)</td>
<td>0.95 (0.92, 0.98)</td>
</tr>
<tr>
<td>Vascularity (n=90)</td>
<td>0.74 (0.60, 0.83)</td>
<td>0.66 (0.48, 0.79)</td>
<td>0.78 (0.66, 0.85)</td>
</tr>
<tr>
<td>Thickness (n=87)*</td>
<td>0.95 (0.92, 0.96)</td>
<td>0.86 (0.78, 0.91)</td>
<td>0.93 (0.89, 0.95)</td>
</tr>
<tr>
<td><strong>‘Worst’ scar site</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation (n=90)</td>
<td>0.96 (0.94, 0.98)</td>
<td>0.95 (0.93, 0.97)</td>
<td>0.97 (0.95, 0.98)</td>
</tr>
<tr>
<td>Vascularity (n=90)</td>
<td>0.84 (0.76, 0.89)</td>
<td>0.67 (0.50, 0.78)</td>
<td>0.73 (0.59, 0.82)</td>
</tr>
<tr>
<td>Thickness (n=57)**</td>
<td>0.95 (0.91, 0.97)</td>
<td>0.91 (0.85, 0.95)</td>
<td>0.96 (0.92, 0.97)</td>
</tr>
<tr>
<td><strong>Normal skin site</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation (n=90)</td>
<td>0.95 (0.93, 0.97)</td>
<td>0.92 (0.87, 0.95)</td>
<td>0.91 (0.86, 0.94)</td>
</tr>
<tr>
<td>Vascularity (n=90)</td>
<td>0.73 (0.58, 0.82)</td>
<td>0.73 (0.58, 0.82)</td>
<td>0.54 (0.30, 0.69)</td>
</tr>
<tr>
<td>Thickness (n=90)</td>
<td>0.94 (0.91, 0.96)</td>
<td>0.92 (0.88, 0.95)</td>
<td>0.95 (0.92, 0.97)</td>
</tr>
</tbody>
</table>

† ICC (95%CI): Intra-class Correlation Coefficient (95% Confidence Interval)
‡ Raters 1-3
* 3 site measurements for one subject were not included in the analysis (reached maximum measurement)
** 33 site measurements not included in the analysis (reached maximum measurement)

<table>
<thead>
<tr>
<th>Component</th>
<th>R1 vs R2‡</th>
<th>R1 vs R3‡</th>
<th>R2 vs R3‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC (95%CI)†</td>
<td>ICC (95%CI)†</td>
<td>ICC (95%CI)†</td>
</tr>
<tr>
<td><strong>‘Best’ scar site</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation (n=30)</td>
<td>0.98 (0.96, 0.99)</td>
<td>0.94 (0.88, 0.97)</td>
<td>0.95 (0.92, 0.98)</td>
</tr>
<tr>
<td>Vascularity (n=30)</td>
<td>0.77 (0.52, 0.89)</td>
<td>0.68 (0.33, 0.85)</td>
<td>0.79 (0.56, 0.90)</td>
</tr>
<tr>
<td>Thickness (n=29)*</td>
<td>0.97 (0.94, 0.99)</td>
<td>0.90 (0.77, 0.95)</td>
<td>0.95 (0.88, 0.98)</td>
</tr>
<tr>
<td><strong>‘Worst’ scar site</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation (n=30)</td>
<td>0.97 (0.94, 0.99)</td>
<td>0.97 (0.93, 0.98)</td>
<td>0.97 (0.95, 0.99)</td>
</tr>
<tr>
<td>Vascularity (n=30)</td>
<td>0.85 (0.69, 0.93)</td>
<td>0.68 (0.32, 0.85)</td>
<td>0.73 (0.43, 0.87)</td>
</tr>
<tr>
<td>Thickness (n=19)**</td>
<td>0.98 (0.96, 0.99)</td>
<td>0.97 (0.91, 0.98)</td>
<td>0.98 (0.94, 0.99)</td>
</tr>
<tr>
<td><strong>Normal skin site</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation (n=30)</td>
<td>0.96 (0.92, 0.98)</td>
<td>0.93 (0.84, 0.96)</td>
<td>0.92 (0.82, 0.96)</td>
</tr>
<tr>
<td>Vascularity (n=30)</td>
<td>0.72 (0.42, 0.87)</td>
<td>0.74 (0.44, 0.87)</td>
<td>0.57 (0.10, 0.79)</td>
</tr>
<tr>
<td>Thickness (n=30)</td>
<td>0.97 (0.93, 0.98)</td>
<td>0.96 (0.92, 0.98)</td>
<td>0.98 (0.97, 0.99)</td>
</tr>
</tbody>
</table>

† ICC (95%CI): Intra-class Correlation Coefficient (95% Confidence Interval)
‡ Raters 1-3
* 1 subject not included in the analysis (reached maximum measurement)
** 11 subjects not included in the analysis (reached maximum measurement)
3.5.3 Test-retest reliability

Repeat scar assessments were completed for 19 out of the 30 subjects. For measurement of pigmentation, vascularity and thickness, 12 subjects had repeat testing conducted by three raters while seven were assessed by only two raters, resulting in a total of 50 repeat tests for analysis. As per protocol, each rater took measurements at three sites of each area of assessment (‘best’ and ‘worst’ area of index scar and the contralateral normal skin area). For analysis of individual measurements (three sites per area in 50 repeat tests), up to 150 paired data sets were available for test-retest analysis. All subjects confirmed to the rater team that nothing significant had affected their scar since the first assessment including no different topical preparations, pressure therapy or excessive sun exposure.

The results of the test-retest reliability ICC and 95% CI for pigmentation, vascularity and thickness are presented in Table 3-4. The number of data pairs included in each analysis is indicated.

Pigmentation had ‘excellent’ test-retest reliability in the ‘best’ and ‘worst’ areas of the scar and the contralateral normal skin area (ICC 0.87, 0.89 and 0.83 respectively).

Vascularity demonstrated varying levels of test-retest reliability. In the ‘worst’ area of the scar test-retest reliability for vascularity was ‘good’ (ICC 0.42). In the ‘best’ area of scar and the contralateral normal skin area test-retest reliability was ‘marginal’ with ICC 0.29 and 0.39 respectively.

Thickness also showed ‘excellent’ test-retest reliability in all the three areas: the ‘best’ area of the scar (ICC 0.97), ‘worst’ area of the scar (ICC 0.92) and the contralateral normal skin area (ICC 0.86).

For pliability, difficulties in obtaining measurements were experienced which limited the number of test-retest data pairs for analysis. Matched measurements (test-retest) were not successfully obtained for six of the 19 subjects at any of the three sites in the ‘best’ area of the scar and the normal skin area. In the ‘worst’ area of the scar, matched measurements for test-retest analysis were not successfully obtained for 10 of the 19 subjects at any of the three sites. Particular difficulty in obtaining successful measurements was experienced in the ‘worst’ area of the scar, on the legs and body areas with thick hair. Results of test-retest reliability for pliability using available matched data were ‘excellent’ in the ‘best’ (ICC 0.91 and 95%CI: 0.78-0.96) and ‘worst’ (ICC 0.76 and 95%CI: 0.29-0.92) areas of scar. However, in the contralateral normal skin area pliability only achieved ‘good’ test-retest reliability (ICC 0.45 and 95%CI: 0.30-0.76).
Table 3-4: Test-retest reliability for pigmentation, vascularity and thickness components using individual site measurements

<table>
<thead>
<tr>
<th>Measurement site</th>
<th>Pigmentation ICC (95% CI)†</th>
<th>Vascularity ICC (95% CI)†</th>
<th>Thickness ICC (95% CI)†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n)‡</td>
<td>(n)‡</td>
<td>(n)‡</td>
</tr>
<tr>
<td>‘Best’ site of scar</td>
<td>0.87 (0.82, 0.90) (n=150)</td>
<td>0.29 (0.01, 0.48) (n=150)</td>
<td>0.97 (0.89, 0.94) (n=150)</td>
</tr>
<tr>
<td>‘Worst’ site of scar</td>
<td>0.89 (0.85, 0.92) (n=150)</td>
<td>0.42 (0.19, 0.58) (n=150)</td>
<td>0.92 (0.88, 0.95) (n=99)*</td>
</tr>
<tr>
<td>Normal skin site</td>
<td>0.83 (0.78, 0.88) (n=150)</td>
<td>0.39 (0.15, 0.56) (n=150)</td>
<td>0.86 (0.81, 0.89) (n=150)</td>
</tr>
</tbody>
</table>

† ICC (95% CI): Intra-class Correlation Coefficient (95% Confidence Interval)
‡ n = number of test-retest data pairs included in analysis
* 51 test-retest data pairs not included in the analysis (at least one measurement reached maximum limit)

3.6 Discussion

The main purpose of this study was to measure scar components analogous to those measured in the subjective VSS, namely, pigmentation (melanin), vascularity (erythema), pliability (elasticity) and height (thickness), using the DermaLab Combo® device. These objective measurements were analysed to assess the inter-rater and test-retest reliability and the capability of the DermaLab Combo® device in conducting burn scar assessments.

The DermaLab Combo® is a relatively new commercially-available integrated skin testing device primarily designed for skin testing in the cosmetic field. The advantages of this device are that it is user-friendly and measures multiple components, including those assessed in the VSS. Therefore, there is potential for this device to be used for clinical and research purposes if it is shown to be reliable in measuring scar components.

The narrow measuring head aperture of the DermaLab Combo® may be a source of potential bias when measuring large scars. A series of measurements are needed to obtain an average score representative of the entire scar (Kaartinen 2011). The current study recorded and used three individual site measurements within the 3 cm x 3 cm areas representing the ‘best’ and ‘worst’ areas of the index scar and contralateral normal skin. The individual site measurements and average of the three site measurements were both used for the purpose of the analysis.

The results of this study show ‘excellent’ inter-rater and test-retest reliability for DermaLab Combo® measurements of pigmentation. This is a marked improvement compared to the mVSS assessment of pigmentation in our previous study (Gankande et al. 2013) that demonstrated only ‘marginal to good’ inter-rater reliability ($k_w 0.02-0.33$) in the ‘worst’ area of scar and ‘good’
inter-rater reliability ($k_w$ 0.45-0.57) in the ‘best’ area of scar. No technical issues were encountered making pigmentation measurements.

Vascularity assessment indicated ‘good’ to ‘excellent’ inter-rater reliability in all scar areas. This represents a marginal improvement compared to the inter-rater reliability of the mVSS reported for the ‘best’ area of the scar in our previous study; ‘good’ to ‘excellent’ inter-rater reliability ($k_w$ 0.64-0.76) in the 'worst' area of the scar, and ‘good’ inter-rater reliability ($k_w$ 0.44-0.71) in the ‘best’ area of the scar (Gankande et al. 2013). However, test-retest reliability for the vascularity component using the DermaLab Combo® was low and failed to achieve an acceptable level of reliability.

The inter-rater reliability of the vascularity measurements was not as high as we had expected, and test-retest reliability was particularly poor. This was in sharp contrast to the ‘excellent’ reliability of the pigmentation measurements. Spectrophotometry, the measuring principle used in the DermaLab Combo® has been in use for over 50 years and is considered to be a reliable and objective method for skin colour assessment (Stamatas et al. 2004; Takiwaki & Serup 1994). Compared to the human observer, reflectance spectrophotometry can detect very small changes in vascularity (erythema) or pigmentation (Latreille et al. 2007). This sensitivity may have been responsible for the lower than expected reliability for vascularity. The protocol of the current study did not specify with what pressure the probe should be applied and excessive pressure may cause blanching of the skin. Conversely, the measurement of pliability with the suction probe within a short space of time before erythema measurement may have increased the degree of erythema. These sources of variation require investigation in future studies, and could be avoided by modifications to the measurement protocol. Therefore, we recommend that protocols of future studies of scar assessment use the colour probe first followed by the ultrasound probe and finally the pliability probe to avoid possible influence on the other measurements. In narrow-band spectrophotometry, changes in the skin chromophore concentrations (melanin and haemoglobin) induce changes in both the melanin index (MI) and the erythema index (EI) in narrow-band spectrophotometry, making it difficult to separate the contribution of each component (Stamatas et al. 2004). This may cause problems for the measurement of skin pigmentation and vascularity after an intervention (e.g. after suction), or over time (e.g. test-retest) where one or both skin chromophores on a location may be altered (Stamatas et al. 2004). Calibration of the colour probe was repeated each day of testing as recommended by the manufacturer, but any issues in the calibration process would also manifest as poor test-retest reliability.

Thickness measurements achieved ‘excellent’ inter-rater and test-retest reliability across all measured areas. Van der Kerckhove and others (2005) used a DermaScan C® (high frequency ultrasound, Cortex Technologies, Denmark) (Cortex Technologies 2011) to measure scar
thickness, and obtained ‘excellent’ results for inter-rater reliability (ICC 0.88). The DermaLab Combo® device shows superior inter-rater reliability for measuring scar thickness compared to the subjective mVSS which demonstrated only ‘good’ to ‘excellent’ inter-rater reliability in both the ‘worst’ and ‘best’ areas of the scar ($k_w$ 0.72-0.76) (Gankande et al. 2013).

Despite achieving ‘excellent’ reliability (both inter-rater and test-retest) of the thickness measurement, a technical limitation was encountered regarding the maximum thickness measurement using the DermaLab Combo®. The manufacturer’s specifications for the DermaLab Combo® ultrasound probe indicate a 3.4 mm penetration capacity. However, the maximum thickness that could be measured during our study was 2.5 mm. Approximately one third of the 30 study subjects had readings in the ‘worst’ area of the scar that reached the maximum measurement. The authors understand an additional ultrasound probe is currently in development which may address this limitation.

The test-retest reliability of pliability using the DermaLab Combo® was ‘excellent’ in the scar areas, but only ‘good’ in normal skin. In this study we did not assess inter-rater reliability of pliability, but Anthonissen and others (2012) using a similar device, the DermaLab® (Cortex Technologies, Denmark) found that inter-rater reliability of pliability of burn scars to be ‘excellent’ for both scars and normal skin (grafted scar: ICC 0.86; spontaneously healed skin and normal skin: ICC 0.93), similar to the inter-rater reliability of pliability measurements made with the mVSS (‘best' area of scar: $k_w$ 0.82-0.84; ‘worst' area of scar: $k_w$ 0.77-0.86) (Gankande et al. 2013). A direct comparison between the DermaLab Combo® and the DermaLab® studies is not possible because the two studies measured different types of reliability and the devices were not identical.

Despite the ‘excellent’ test-retest reliability of pliability measurements with the elasticity (pliability) probe in scar areas, the probe had limitations. The raters found scar assessment with the probe was not achievable on approximately half of the ‘worst’ scar areas, suggesting the probe may not have appropriate specifications for making measurements on rigid tissue. Anthonissen and others (2012) also discuss the potential limitations of the elasticity (pliability) probe of the DermaLab® device to measure rigid scars. Difficulties were also observed obtaining successful measurements on the legs and we speculate that this may be due to the greater thickness of the epidermis of the leg skin. In addition, the probe was ineffective in obtaining a proper grip with the skin with its adhesive tape when the subjects had a rough growth of body hair. The influence of body hair may help explain the low test-retest reliability for normal skin, as normal skin has more hair than scar. Protocols of future studies should include shaving the area prior to use of the elasticity probe.

A major strength of this study is the identification and systematic evaluation of a device with the potential capability of measuring four scar components considered important in scar assessment
in an objective manner. This study has translated the DermaLab Combo® device, currently used in the cosmetic industry, to the area of burn scar assessment with some success. Pigmentation measurement was highly reliable, showing a substantial improvement compared to the mVSS assessment of pigmentation, and no technical issues were encountered. Scar thickness measurement was also highly reliable; however, for about one third of the scars measured the thickness reached the maximum capacity of the device in its current version. Pliability measurement offered only a modest improvement in inter-rater reliability compared to the mVSS and was limited in its capacity to make measurements in approximately half of the burns scars tested. We are of the view that the vascularity assessment using the DermaLab Combo® did not reach its full potential in the current study.

### 3.7 Conclusion

The DermaLab Combo® is an easy to use and commercially available device, making it a viable option for scar assessment in both clinical and research settings. It provides objective and reliable measurements for pigmentation, thickness and pliability. However, the device has limitations in making measurements of thickness and pliability in some burns scars. Thickness can be measured in all scars but reaches a maximum measurement. The problems encountered obtaining successful measurements of pliability and reliable vascularity measurements are being examined further. If future studies provide protocols to improve test-retest reliability of vascularity measurements and obtain pliability measurements more successfully, the DermaLab Combo® will be a valuable option for scar assessment and monitoring.
Chapter 4
Chapter 4 – Exploring the interpretation of the DermaLab Combo® pigmentation and vascularity measurements via the mVSS

This chapter describes the clinical study conducted to explore the relationship between DermaLab Combo® continuous measurements of pigmentation and vascularity of burns scars and the mVSS and obtain evidence to support validity of DermaLab Combo® pigmentation and vascularity measurements. This chapter is based on the following original article titled ‘Interpretation of the DermaLab Combo® pigmentation and vascularity measurements in burn scar assessment: an exploratory analysis’ accepted for publication on 20-01-2015 in Burns prior to the submission of this thesis (Gankande et al. 2015) [Appendix C].

4.1 Introduction

Scar assessment is essential for diagnosis, monitoring, evaluation and therapeutic management of scar pathology (Bayat et al. 2003). Colour evaluation has remained problematic even with advances in the area of scar assessment scales (Kaartinen 2011; Fearmonti et al. 2010). Inter-rater reliability of subjective scales in evaluating colour has been shown to be poor (Kaartinen 2011; Brusselaers et al. 2010a) and requires multiple observers to achieve reliable assessment (Kaartinen 2011; Van Zuijlen et al. 2002). In the clinical setting, scars are typically assessed by a single observer and frequently by many different observers over time, further weakening the accuracy of subjective assessment.

The VSS is a commonly used scar assessment method in many burns care facilities because it is inexpensive and relatively easy to perform. The VSS is a subjective method and measures four scar parameters: pigmentation, vascularity, pliability and height. A modified VSS (Baryza & Baryza 1995) is currently used at the RPH in WA to conduct scar assessments. In a previous study the inter-rater reliability of this method was established (Gankande et al. 2013). For pigmentation inter-rater reliability was ‘marginal’ in both ‘best’ (kω 0.5-0.6) and ‘worst’ (kω 0.1-0.3) areas of the scar while for vascularity inter-rater reliability was ‘good to excellent’ (in both ‘best’ [kω 0.4-0.7] and ‘worst’ [kω 0.6-0.8] areas the scar). This study made recommendations for improving the inter-rater reliability of pigmentation and vascularity components by using more objective tests (Gankande et al. 2013).

The DermaLab Combo® (Cortex Technologies, Denmark) is a commercially-available skin testing device with the capability to measure scar colour (pigmentation and vascularity), pliability and thickness. It was primarily designed for skin testing in the cosmetic field and does
not claim to be a medical device. However, it is a high-specification device based on proven technologies. The DermaLab Combo® measures skin colour by narrow-band spectrophotometry and provides continuous scar measurements for pigmentation (melanin index) and vascularity (erythema index). Our previous work compared the reliability of subjective scar assessments of the mVSS and objective measurements made with the DermaLab Combo®. The results showed that the DermaLab Combo® measures pigmentation and vascularity of a scar more reliably than the mVSS (Gankande et al. 2014).

For a diagnostic tool to have clinical utility it must demonstrate acceptable validity and reliability. Validity refers to evidence that supports that an instrument or tool is actually measuring what it says it is measuring (Campbell & Machin 1993). Concurrent validity in the context of diagnostic instruments, is tested by comparing results obtained with a ‘test’ instrument with those of a ‘gold standard’ (Campbell & Machin 1993; Christie et al. 1997). However, a ‘gold standard’ assessment tool is currently not available for assessing pigmentation and vascularity of scars (Idriss & Maibach 2009). In the absence of a ‘gold standard’ for validation of the DermaLab Combo® device this study used the mVSS (Baryza & Baryza 1995) as a reference standard. The study was conducted at the RPH burns outpatient clinic, Western Australia. This study aims firstly, to examine how the DermaLab Combo® continuous measurements of pigmentation and vascularity of burns scars relate to the mVSS, the current standard clinical scar assessment method used in many burns units; and secondly, to obtain evidence to support the concurrent validity of DermaLab Combo® measurements for pigmentation and vascularity.

4.2 Method

4.2.1 Subjects

Adult patients scheduled for regular scar assessments at the RPH burns outpatient clinic as part of routine follow-up care for a burn injury were targeted for recruitment. Participation was voluntary and all patients received the same standard of clinical care whether they elected to participate or not. The subjects were included in the study if they met the following inclusion criteria: at least 18 years of age; able to provide voluntary written informed consent; had a fully epithelialised burn scar with an area of at least 3 cm x 3 cm; and an adjacent normal skin area of at least 3 cm x 3 cm with similar sun exposure. Subjects were excluded from the study if they were unable to provide informed written consent. All eligible subjects were provided with a patient information sheet and a verbal explanation of the study and written consent was sought.
4.2.2 Study design

The study was a single arm observational study with three scar raters. The three raters were independent and blinded to the results of each other. The study was approved by RPH and UWA Human Research Ethics Committees.

4.2.3 Raters

The three raters used in the study were all experienced scar assessors and included a Senior Occupational Therapist (OT) with over seven years’ experience in the mVSS method (Stratifying rater – R1), a clinical researcher with over two years’ experience in scar assessment and in the use of the DermaLab Combo® device (DermaLab Combo® rater – R2), and a clinician expert rater with over 10 years’ experience in the mVSS method (Burns Surgeon–R3). For the remainder of this paper, raters will be referred to as R1, R2 and R3.

4.3 Scar assessment

4.3.1 The mVSS (Baryza and Baryza modification)

The mVSS (Baryza & Baryza 1995) was used to assess pigmentation and vascularity of the index scar area and the adjacent normal skin area. This mVSS methodology applied has been fully described in our previous publication (Gankande et al. 2013).

4.3.2 The DermaLab Combo® device – colour probe

The colour measurement of the DermaLab Combo® device (Figure 4-1) is based on the principle of narrow-band reflectance spectrophotometry (550 nm ± 30 nm and 660 nm ±60 nm for haemoglobin [erythema or vascularity] and melanin [pigmentation] respectively). The colour probe specifications has been fully described in our previous publication (Gankande et al. 2014). The unit displays four readings (three individual measurements and the average of the three measurements) separately for erythema (vasularity) and melanin (pigmentation) using Commission Internationale de l'Eclairage (CIE) – luminance (L), red-green axis (a) and blue-yellow axis (b) [CIELab] values.
4.4 Data collection and procedure

Data was collected over a four month period, from November 2013 to February 2014. Each rater was blinded to the results of the other raters and independently recorded results in scar assessment forms. All scar assessment forms were filed separately until the data collection was complete.

Study subjects were exposed to constant environmental conditions at the time of the assessments and all assessments were performed with the subjects in a sitting position. The anatomical body part containing the scar was placed in a nondependent position; if the scar was on a leg – both legs were elevated to rest on another chair, if the scar was on an arm – both arms were rested parallel to each other on a table. All pressure garments and bandages were removed at least 15 minutes prior to the initial assessment. A minimum 10 minute wash-out period was observed between each rater. Each subject had three scar assessments performed within one hour on the same day by all three raters.

The stratifying rater (R1) identified and marked with a semi-permanent marker, 3 cm x 3 cm areas on both the index scar (based on the mVSS) and an adjacent normal skin area (Tyack et al.
Subjects were stratified by pigmentation categories (normal pigmentation, hypopigmentation, mixed pigmentation and hyperpigmentation) and vascularity categories (normal vascularity, pink, red and purple) during recruitment until each category achieved a minimum of 20 subjects. The data was recorded in the mVSS data collection sheet (Gankande et al. 2013) and the data stratification sheet (Appendix 4).

The DermaLab Combo® scar assessment (R2) was conducted on the marked index scar area and the adjacent normal skin area using the DermaLab Combo® colour probe. Within the index scar, R2 assessed three individual sites for pigmentation and vascularity (Figure 4-2). The three measurements and the automatically generated average measurement was recorded on a scar assessment form (Appendix 5). Figure 4-2 shows the probe placement for the DermaLab Combo® measurements within both the two 3 cm x 3 cm measurement areas (index scar and the adjacent normal skin). The circles numbered 1, 2 and 3 represent the locations for each site measurement (probe placement). The locations themselves were not marked, as this would interfere with the readings.

Pigmentation and vascularity parameters were scored by clinical expert rater (R3) using the mVSS for the marked index scar and the scores were recorded in the scar assessment form (Appendix 5).

![Figure 4-2: Probe placement for the DermaLab Combo® measurement within the 3 cm x 3 cm scar and the normal skin areas](image)

### 4.5 Analysis and sample size

Descriptive statistics for age, gender, Fitzpatrick skin type and anatomical location of scar were performed. The DermaLab Combo®-generated average measurements were used to derive study variables for scar pigmentation and vascularity relative to the normal skin measurements: melanin index % (MI%) (pigmentation) and erythema index % (EI%) (vascularity).
4.5.1 Generation of melanin index % (MI%) and erythema index % (EI%)

After personal communication with the manufacturer of the DermaLab Combo® relative percentage changes in melanin and erythema were quantified. The percentage change in pigmentation (ΔP) was calculated as the difference between the DermaLab Combo® pigmentation measurements for scar (Ps) and normal skin (Pn) divided by the normal skin measurements (Pn) and multiplied by 100 (ΔP = \( \frac{(Ps-Pn)}{Pn} \times 100 \) [%]). The following formula was used to generate the variable, melanin index % (MI%) with normal skin pigmentation equated to 100%: MI% = 100 + ΔP (%). Similarly, the percentage change in vascularity (ΔV) was calculated as the difference between the DermaLab Combo® vascularity measurement for the scar (Vs) and the normal skin (Vn) divided by the normal skin measurement (Vn) and multiplied by 100 (ΔV = \( \frac{(Vs-Vn)}{Vn} \times 100 \) [%]). The following formula was then used to generate the variable, erythema index % (EI%), with normal skin vascularity equated to 100%: EI% = 100 + ΔV (%).

4.5.2 Exploratory data analysis

Data was graphed and visualised, descriptive statistics generated (median; IQR; minimum and maximum) and outliers identified for the variables MI% and EI% across the mVSS categories of pigmentation (normal pigmentation, hypopigmentation and hyperpigmentation) and vascularity (normal, pink, red, purple). Kruskal-Wallis test with post-test comparisons for non-parametric distributions were conducted. Dot plots with summary measures (median, minimum – maximum) were generated for pigmentation MI% and vascularity EI% by clinical classification using the mVSS.

To explore if the two variables were more useful than a single variable for discriminating mVSS categories, bivariate plots were generated and analyses of MI% and EI% data by mVSS categories were undertaken (Balakrishnama & Ganapathiraju 1999; Li et al. 2011). A linear discriminant axis that optimised separation of mVSS categories was identified using an iterative process from the bivariate plot. A discriminant score for each observation was calculated from both variables using the formula of the line perpendicular to the discriminant axis which passes through the observation: discriminant score (pigmentation) = \( a_1 \) MI% + \( b_1 \) EI%; discriminant score (vascularity) = \( a_2 \) EI% + \( b_2 \) MI%. Cut-off discriminant values for classification into mVSS categories were identified. An iterative process was used to identify discriminant cut-offs that optimised sensitivity and specificity. The concordance of classification into mVSS...
categories using two predictor variables was compared to results from using cut-off values for the single variable (MI% or EI%).

To assess the associations between the MI% and EI% derived values from the DermaLab Combo® measurements (continuous) with the respective coded numerical scores of mVSS (ordinal) pigmentation and vascularity, Kendall tau-b (τ_b) rank correlations (accounting for ties) were performed (Khamis 2008).

In the absence of a ‘gold standard’, the subjective mVSS classification was used as the reference standard and analyses were restricted to the data where there was absolute agreement between R1 and R3.

A sample size of 100 subjects was recruited to ensure a minimum of 20 subjects in each of the mVSS categories of pigmentation and vascularity to conduct a meaningful exploratory analysis (Microsoft MedCalc version 12.7.0 2013). This sample size was also considered adequate to conduct Receiver Operating Characteristic (ROC) analysis, if feasible, with level of significance 0.05 and 80% power to detect a minimum ROC area under curve of 0.75 (Microsoft MedCalc version 12.7.0 2013). After recruitment it was decided that the mVSS classification category of mixed pigmentation would be excluded from the analysis due to the heterogeneous nature of mixed pigmented scars. A scar classified as mixed pigmentation by mVSS classification may include areas of hypopigmentation, normal pigmentation and hyperpigmentation and requires a different scar sampling protocol to that used in this study. There were insufficient data to perform ROC analyses. All data analyses were performed using Stata V12 statistical software (StataCorp LP, College Station, Tex).

4.6 Results

4.6.1 Subject stratification and descriptive statistics

Of the 100 subjects recruited, 61% were male and the median age was 35 years (IQR: 26-52 years; minimum-maximum: 18-73 years). The anatomical locations of the assessed scars were: arm (n=44), leg (n=45), chest and abdomen (n=6), head and neck (n=1), groin and buttock (n=1) and back (n=3). The Fitzpatrick skin type (identified from outpatient medical records for 85 out of 100 subjects) for the subjects were: Type II (n=33), Type III (n=41), Type IV (n=7) and Type V (n=4). Final figures of stratification at recruitment are presented in Table 4-1.
Table 4-1: The number of subjects recruited using the mVSS (R1) and absolute agreement between R1 and R3 used in the analyses

<table>
<thead>
<tr>
<th>mVSS Category</th>
<th>Subjects recruited using mVSS (R1)</th>
<th>Absolute agreement between mVSS raters (R1 and R3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Pigmentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal pigmentation</td>
<td>27</td>
<td>19</td>
</tr>
<tr>
<td>Hypopigmentation</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Hyperpigmentation</td>
<td>29</td>
<td>19</td>
</tr>
<tr>
<td>Total*</td>
<td>73</td>
<td>54</td>
</tr>
<tr>
<td>Vascularity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>Pink</td>
<td>28</td>
<td>16</td>
</tr>
<tr>
<td>Red</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td>Purple</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>66</td>
</tr>
</tbody>
</table>

*Mixed pigmentation category was excluded from analysis (27 subjects; 8 disagreements R1 vs. R3)

4.6.2 The mVSS rater agreement

There was disagreement in mVSS scar classification between R1 and R3 for 27/100 study subjects for pigmentation, and 34/100 study subjects for vascularity. Table 4-1 presents mVSS pigmentation and vascularity classifications for the total number of subjects recruited and where there was absolute agreement between R1 and R3.

4.6.3 Exploratory data analysis – pigmentation

Results of univariate descriptive statistics for the melanin index percent (MI%) for mVSS categories of pigmentation (normal pigmentation, hypopigmentation and hyperpigmentation) are presented in Table 4-2. The Kruskal-Wallis test with post-test comparisons indicated statistically significant differences in distribution between the following pigmentation categories ($\chi^2 16.84; p = 0.004$; degrees of freedom: 3, n=54): hypopigmentation vs normal pigmentation and hypopigmentation vs hyperpigmentation; no statistically significant differences in MI% for normal pigmentation vs hyperpigmentation were found. Kendall $\tau_b$ correlation analyses found statistically significant positive associations of moderate strength between the DermaLab Combo® MI% values for pigmentation and R1 mVSS ($\tau_b 0.4, p<0.001$) and R3 mVSS ($\tau_b 0.4, p<0.001$) pigmentation scores. Three outliers of the normal pigmentation mVSS category were identified within the hyperpigmentation category and excluded from the analyses. These
subjects had high vascularity readings by both the mVSS and the DermaLab Combo® (EI%), and scars may have been misclassified as normally-pigmented instead of hyperpigmented. In the presence of a high vascularity score it is difficult to assess the ‘brownness’ of the scar.

Table 4-2: Descriptive statistics for MI% values for mVSS pigmentation categories

<table>
<thead>
<tr>
<th>mVSS Pigmentation Category</th>
<th>N</th>
<th>Median (IQR)*</th>
<th>Min†</th>
<th>Max†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypo (1)</td>
<td>16</td>
<td>94.1 (69.2-120.0)</td>
<td>69.2</td>
<td>120.0</td>
</tr>
<tr>
<td>Normal (0)</td>
<td>19</td>
<td>110.0 (70.6-175.0)</td>
<td>70.6</td>
<td>175.0</td>
</tr>
<tr>
<td>Hyper (2)</td>
<td>19</td>
<td>127.7 (74.2-240.7)</td>
<td>74.2</td>
<td>240.7</td>
</tr>
</tbody>
</table>

* IQR= inter quartile range
† Min-Max= minimum-maximum
Pairwise comparisons for MI% (Normal pigmentation vs. Hypopigmentation p=0.003; Normal pigmentation vs. Hyperpigmentation p=0.04; Hypopigmentation vs. Hyperpigmentation p<0.001)

Dot plots for MI% scores for mVSS pigmentation categories normal, hypopigmentation and hyperpigmentation are presented in Figure 4-3. Half of the MI% values for scars classified as hypopigmented (mVSS) demonstrated a decrease in the MI% value relative to normal pigmentation (i.e. <100%). For scars classified as normal pigmentation by the mVSS, 8/19 had an MI% value close to normal (MI% 100 ± 10%) with almost half of the MI% values demonstrating an increase in pigmentation of 10% or greater relative to normal (MI% >110%); for two cases, results indicated a decrease in pigmentation of greater than 10%. For scars classified as hyperpigmented by the mVSS, the majority of the MI% values showed a greater than 10% increase in pigmentation relative to normal (MI% >110%). However, for two cases MI% values were either close to normal (MI% 100 ± 10%) or showed a decrease in pigmentation of greater than 10% (MI% <90%).
The mVSS classification of the DermaLab Combo® results using univariate and bivariate approaches are shown in Figure 4-4 and concordance results are summarised in Table 4-3. Bivariate plots of the MI% and EI% scores by mVSS pigmentation categories are presented in Figure 4-4. Figure 4-4A shows univariate pigmentation classification (MI%) using cut-off values identified (<90%: hypopigmentation; 90-125%: normal pigmentation; and >125%: hyperpigmentation). Classification concordance was 60% (Table 4-3A) and increased by 2% with omission of three outliers from the analysis (Table 4-3B). Figure 4-4B presents bivariate classification of pigmentation using discriminant cut-off values (<208: hypopigmented; 208-257: normal; >257: normal; >257: hyperpigmented). Using these discriminant cut-off values for pigmentation, 74% concordance was achieved (Table 4-3C); concordance was increased by 6% (80%) with omission of three outliers from analysis (Table 4-3D).

Classification of the data by mVSS pigmentation categories using two predictor variables (MI% and EI%) rather than one variable (MI%) resulted in improvement in classification concordance of 13% and 17%, for analyses using all values and values omitting outliers, respectively (Table 4-3).
Figure 4-4: MI% and EI% values by mVSS pigmentation category (A) Cut-off values on the X axis (MI%) (B) Cut-off values on the transformed axes
### Table 4-3: Classification concordance results for MI% by mVSS pigmentation categories for cut-off values using standard and transformed X axis

#### Cut-off values with x axis (MI%)

<table>
<thead>
<tr>
<th>MI%</th>
<th>Hypo</th>
<th>Normal</th>
<th>Hyper</th>
<th>Total</th>
<th>MI%</th>
<th>Hypo</th>
<th>Normal</th>
<th>Hyper</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>mVSS</td>
<td>&lt;90</td>
<td>100-125</td>
<td>&gt;125</td>
<td></td>
<td>mVSS</td>
<td>&lt;90</td>
<td>100-125</td>
<td>&gt;125</td>
<td></td>
</tr>
<tr>
<td>Hypo</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>11</td>
<td>Hypo</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Normal</td>
<td>1</td>
<td>8</td>
<td>5</td>
<td>14</td>
<td>Normal</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Hyper</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>13</td>
<td>Hyper</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>17</td>
<td>14</td>
<td>38</td>
<td>60.5%</td>
<td>Total</td>
<td>7</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Cut-off values with transformed axis

<table>
<thead>
<tr>
<th>MI%</th>
<th>Hypo</th>
<th>Normal</th>
<th>Hyper</th>
<th>Total</th>
<th>MI%</th>
<th>Hypo</th>
<th>Normal</th>
<th>Hyper</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>mVSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mVSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypo</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>Hypo</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Normal</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>14</td>
<td>Normal</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Hyper</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>13</td>
<td>Hyper</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>9</td>
<td>15</td>
<td>38</td>
<td>73.7%</td>
<td>Total</td>
<td>14</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Deleted outliers = (124,300) (135,275) (175,410)

#### 4.6.4 Exploratory data analysis – vascularity

Results of descriptive statistics for the variable erythema index percent (EI%) overall and for mVSS categories of vascularity (normal vascularity, pink, red and purple) are presented in Table 4-4. Kruskal-Wallis test with post-test comparisons identified statistically significant differences in EI% medians for vascularity categories of normal vs pink, normal vs red and normal vs purple; no other pairwise comparisons of vascularity demonstrated statistically significant differences [$\chi^2$ (degrees of freedom: 3, n=66) = Chi square: 17.56 (p=0.001)]. Kendall tau_b ($\tau_b$) correlation analyses found statistically significant positive associations of moderate strengths for DermaLab Combo® EI% values for vascularity with R1 mVSS ($\tau_b$ 0.4, p<0.001) and R3 mVSS ($\tau_b$ 0.3, p<0.001) vascularity scores.
Table 4-4: Descriptive statistics for EI% values for mVSS vascularity categories

<table>
<thead>
<tr>
<th>mVSS Vascularity Category</th>
<th>N</th>
<th>Median (IQR)*</th>
<th>Min†</th>
<th>Max†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (0)</td>
<td>22</td>
<td>73.2 (26.5-159.3)</td>
<td>26.5</td>
<td>159.3</td>
</tr>
<tr>
<td>Pink (1)</td>
<td>16</td>
<td>103.8 (58.1-200.0)</td>
<td>58.1</td>
<td>200.0</td>
</tr>
<tr>
<td>Red (2)</td>
<td>16</td>
<td>130.0 (87.5-410.0)</td>
<td>87.5</td>
<td>410.0</td>
</tr>
<tr>
<td>Purple (3)</td>
<td>12</td>
<td>127.1 (63.3-300.0)</td>
<td>63.3</td>
<td>300.0</td>
</tr>
</tbody>
</table>

* IQR= inter quartile range, † Min-Max= minimum-maximum
Pairwise comparisons for EI% (Normal vs Pink p=0.01; Normal vs Red p<0.001; Normal vs Purple p<0.001; Pink vs Red p=0.08; Pink vs Purple p=0.19; Red vs Purple p=0.33)

Figure 4-5A represents the dot plot of EI% values for vascularity categories of normal vascularity, pink, red and purple. For scars classified as normal vascularity by the mVSS, one case had an EI% close to normal (EI% 100 ±10%), 15/22 of EI% values showed a greater than 10% decrease in vascularity relative to normal (EI% <90%) and 6/22 of EI% values showed a greater than 10% increase in vascularity relative to normal (EI% >110%). For scars classified as pink by the mVSS, nearly half of the EI% values (7/16) showed a greater than 10% increase in vascularity relative to normal (EI% >110%). However, 6/16 had an EI% values close to normal (EI% 100 ±10%) and in a small number of scars (3/16) the EI% values indicated a greater than 10% decrease in vascularity relative to normal (EI% <90%). For scars classified as red by the mVSS, 12/16 of EI% values showed a greater than 10% increase in vascularity relative to normal (EI% >110%). The remaining EI% values (4/16) were close to normal (EI% 100 ±10%). The highest EI% score for scars classified as pink by the mVSS was 200%.

Comparing the data in Figures 4-5, there is considerable overlap of EI% value between scars classified as pink and red using the mVSS. For 8/12 scars classified as purple by the mVSS had an EI% value with a greater than 10% increase in vascularity relative to normal (EI% >110%). Three of the scars classified as purple by the mVSS had EI% values close to normal (EI% 100 ±10%) and one scar had an EI% value with greater than 10% decrease in vascularity relative to normal (EI% <90%). Comparison of the distributions of EI% values showed that 11/12 scars classified as purple by the mVSS had EI% scores nested within the range of EI% values for scars classified as red.
Figure 4-5: Dot plot of EI% values of scars for vascularity by mVSS with summary statistics (median, minimum-maximum)

Bivariate plots of the classification of EI% and MI% values by mVSS vascularity categories are presented in Figure 4-6. Visual examination identified no visible clustering of data (univariate (EI%) or bivariate) by mVSS vascularity categories (normal, pink, red and purple) and no further analyses were conducted.

Figure 4-6: EI% and MI% values by mVSS vascularity category
4.7 Discussion

This study has provided a useful starting point and valuable preliminary information in the use of DermaLab Combo® pigmentation and vascularity measurements in burns scar assessment. Our study showed that quantifying the relative percentage changes in erythema (EI%) and melanin (MI%) was a useful way to explore how the DermaLab Combo® pigmentation and vascularity measurements related to mVSS classification. In addition, positive associations of moderate strength were demonstrated between the mVSS clinical classification of pigmentation and vascularity and the DermaLab Combo® derived MI% and EI% values, respectively. However, further investigation is required to obtain evidence to support concurrent validity of the DermaLab Combo® pigmentation and vascularity measurements in burns scar assessment.

The results of this study showed that MI% values as a measure of scar pigmentation did not clearly fit within distinct ranges according to the mVSS pigmentation categories. Less than half of the scars classified as hypopigmentation (43%) and normal pigmentation (42%) using the mVSS had a MI% score that fitted within the appropriate range (MI% <90, MI% 100 ± 10 respectively). However, for the majority (89.5%) of scars classified as hyperpigmented by the mVSS, MI% scores were greater than 110%. The concordance associated with classification of mVSS pigmentation categories using univariate cut-off values of MI% was 63% (Table 4-3). Examination of mVSS pigmentation categorisation using the two variables, MI% and EI%, rather than MI% alone, improved pigmentation categorisation with greater concordance (80%). Changes in the skin chromophore concentrations (melanin and haemoglobin) induce changes in both MI and EI indices in narrow-band reflectance spectrophotometry, making it difficult to separate the contributions of each component (Stamatas et al. 2004) and this may explain the improved concordance of categorisation using the two variables.

The poor fit between the MI% values and the mVSS categories of normal pigmentation and hypopigmentation may be attributed to several factors. The low inter-rater reliability of the pigmentation component of the mVSS means that misclassification of mVSS categories was possible even when the two raters agreed. In an earlier study we found that inter-rater reliability for pigmentation in the mVSS was dependent on the severity of the scar and the combination of raters (Gankande et al. 2013). Another possible factor that may have contributed to the poor fit between the MI% scores and the mVSS categories of normal pigmentation and hypopigmentation is the influence of body hair on the melanin estimates obtained with spectrophotometry (Van der Mei et al. 2002). While the index scar areas of our subjects were hairless, the normal skin areas of our subjects were not shaved, and this may have influenced the MI % scores.
The results of the study showed that EI% values did not fit within ranges that reflected the mVSS vascularity categories. The majority of scars (68%) classified as normal (mVSS) had an EI% values less than normal (EI% <90) and 27% of scars had EI% scores greater than normal (EI% >110). There was considerable overlap between the pink and normal categories. Only 18.7% of red scars had an EI% values greater than 200%, the highest EI% value in the pink category, and over 80% of red scars were nested within the mVSS pink category. Approximately 92% of scars classified as purple had EI% values nested within the mVSS red category. Examination of mVSS vascularity categorisation using two predictor variables, EI% and MI%, did not identify any patterns of clustering of values by mVSS categories (Figure 4-6).

The poor fit between the DermaLab Combo® measurements and the mVSS vascularity categories may reflect the wide range in the degree of agreement between raters (k, 0.4-0.8) reported for the vascularity component of the mVSS (Gankande et al. 2013). Another factor for consideration is that measurement of erythema by narrow-band reflectance spectrophotometry is very sensitive and may detect changes that are not solely due to scar vascularity, and as such, any changes in skin or scar erythema due to touch or palpation may be detected (Reed et al. 1961).

While this study provided an initial examination of the use an objective scar assessment tool, the study had several limitations. The major limitation was the lack of an available ‘gold standard’ in scar assessment and the use of subjective clinical expert opinion as the reference standard (mVSS) to compare DermaLab Combo® measurements. Expert opinion is considered to be a reasonable proxy when validating health instruments in the absence of a ‘gold standard’ (Katz et al. 2000; Sullivan Pepe 2003; Fletcher et al. 1996). Future validation studies of the DermaLab Combo® measurements of pigmentation and vascularity may be possible using standard colour reference cards developed by the cosmetic industry, for example the Skin Color Chart® (L’Oréal) (de Rigal et al. 2007). The sample size for this study was based on sufficient base-line numbers within each mVSS category for both pigmentation and vascularity; however, this did not allow for disagreement between raters (R1 and R3), that resulted in a reduced sample size for analysis. The heterogeneous nature of the study sample may also have introduced a limitation. The inclusion of subjects from a range of skin types, burn scars at different stages of healing (data not shown) and scars in different anatomical locations created heterogeneity within the study sample. Despite the study sample being representative of our general patient population at RPH, it may have confounded our results. Campbell and Machin (1993) discuss that a diagnostic assessment may have variable levels of sensitivity at different stages of a condition, and this may be applicable to our study. For future studies it is suggested that larger sample sizes be used where study subjects are stratified by skin type, stage of healing and the location of scar, to obtain more uniform samples for analyses. Study protocols that include a longer washout period (minimum of 20 minutes) between scar assessments and the shaving of
normal skin areas prior to scar assessment are also recommended. Given the lack of a current ‘gold standard’ measurement tool for scar assessment, further validity studies using the mVSS as standard reference may be improved using a greater number of raters to improve the accuracy of the assessment.

The mVSS mixed pigmentation classification refers to a scar with uneven areas of hypopigmentation, normal pigmentation and hyperpigmentation. A measure of spread of the DermaLab Combo® pigmentation measurement will vary depending on the heterogeneity of the scar pigmentation composition and also the skin type of the individual. To progress the understanding of the DermaLab Combo® measurements in context of the clinical category of mixed pigmentation (mVSS), future studies may consider a protocol that either identifies specific sub-sites of hypo, normal and hyper pigmentation within the heterogeneous pigmented scar for data collection and ongoing surveillance or one that systematically identifies a greater number of sub-sites (e.g. grid) within the mixed pigmentation scar.

The major strength of this study is that it has provided useful preliminary information in the use of DermaLab Combo® pigmentation and vascularity measurements in burns scar assessment. Our study showed that quantifying the relative percentage changes in erythema and melanin, i.e. the increase or decrease in values of scar compared to the normal skin, was a useful starting point in understanding the DermaLab Combo® pigmentation and vascularity measurements. The results for pigmentation using MI% showed clustering of values into mVSS categories was further improved by incorporating EI%. The results for vascularity using EI%, however, did not show any such improvements of clustering of values into mVSS categories by incorporating MI%. This study also provides a useful research platform from which further investigations can be undertaken with respect to strengthening protocols, including sampling strategies, and the need to pursue other means or tools to assess the validity of the DermaLab Combo® measurements that will have clinical utility in burn scar assessment.

4.8 Conclusion

This study provides initial evidence in the interpretation of the DermaLab Combo® measurements in the clinical context of burns scar assessment. Results of our study showed that the MI% values derived from the DermaLab Combo® measurements could be used to differentiate pigmentation categories classified by the mVSS (hypopigmentation, normal pigmentation and hyperpigmentation). The DermaLab Combo® is an objective tool providing continuous numerical data that may be useful for clinical scar monitoring over time; however, further work is required to understand and optimise the interpretation of the DermaLab Combo® measurements for pigmentation and vascularity. Currently, scar assessment at RPH
Burns Unit and other burns care facilities use the subjective VSS or a modified version (i.e. mVSS). Therefore, improvements in the interpretation of the DermaLab Combo® measurements and classification of burn scar parameters analogous to the mVSS will be valuable.
Chapter 5 – Evaluation of the clinical studies, general discussion, conclusions and recommendations

This final chapter presents evaluations of the clinical studies undertaken in this project. Issues related to measurement and the complexities of interpreting the DermaLab Combo® objective scar measurements are discussed. The implications of the projects results are presented, and the strengths and limitations of the work are identified.

5.1 Evaluations of the clinical studies

5.1.1 Reliability of the mVSS (Study I – Chapter 2)

5.1.1.1 Summary of study

Three raters performed scar assessments on thirty patients with burn scars using the mVSS-TBSA. Scoring on pigmentation, vascularity, pliability and height was undertaken for the ‘best’ and ‘worst’ areas of each scar. Raters also allocated the total body surface area of the scar (%TBSA) to three mVSS categories (<5, 5-10, >10). Intra-class correlation coefficient (ICC) and weighted kappa statistic (k_w) and were used to assess inter-rater reliability. Results are summarised in Table 5.1.

<table>
<thead>
<tr>
<th></th>
<th>Inter-rater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pliability</td>
<td>‘excellent’</td>
</tr>
<tr>
<td>Height</td>
<td>‘good’ to ‘excellent’</td>
</tr>
<tr>
<td>Vascularity</td>
<td>‘good’ to ‘excellent’</td>
</tr>
<tr>
<td>Pigmentation</td>
<td>‘marginal’ to ‘good’</td>
</tr>
<tr>
<td>Total mVSS score</td>
<td>‘good’ to ‘excellent’</td>
</tr>
<tr>
<td>Inter-rater reliability of mVSS scores depended on the severity of the scar area being assessed.</td>
<td></td>
</tr>
<tr>
<td>Allocation of scar %TBSA to two broad mVSS categories (&lt;5, ≥5) showed 'good’ to ‘excellent’ inter-rater reliability.</td>
<td></td>
</tr>
</tbody>
</table>

5.1.1.2 Measurement issues

Measurement of pigmentation showed the lowest inter-rater reliability among the four scar parameters measured by the mVSS (k_w 0.02-0.33), similar to the inter-rater reliability reported by Nedelec and others (2000) (k_w 0.09-0.14) and Forbes-Duchart and others (2007) (k_w 0.16-0.21). However, Baryza and Baryza (1995) reported greater inter-rater reliability (k_w 0.61) with
the use of a plexiglass tool to remove the colour due to vascularity by ‘blanching’ the skin. While the study protocol in this thesis specified the use of the plexiglass tool, the other studies cited here do not mention if the plexiglass tool was used. The results showed that further improvements in reliability are required; a protocol specifying the amount of pressure to be placed on the plexiglass tool by each rater, and monitoring rater compliance may ensure more consistent removal of colour due to vascularity.

The inter-rater reliability of vascularity measurements was rated as ‘good’ ($k_w$ 0.64-0.76), similar to the original Baryza and Baryza (1995) ($k_w$ 0.73) and considerably greater than the inter-rater reliability reported by Nedelec and others (2000) ($k_w$ 0.14-0.25) and Forbes-Duchart and others (2007) ($k_w$ 0.04-0.25). The influence of physical factors (e.g. touch or palpation, body position of the anatomical site containing the scar) and environmental conditions (e.g. temperature of the examination room) need consideration. In order to minimise physical influences on vascularity measurements, strict control of environmental conditions and the order of measurements during scar assessment is essential. The mVSS protocols need to specify the following: vascularity should be the first scar component assessed; a wash-out period of at least 20 minutes between repeat scar assessments needs to be maintained; the temperature of the examination room needs to be specified and kept constant; and, the anatomical body part containing the scar should be positioned in a non-gravity dependent manner for at least 15 minutes prior to assessment. Other authors have trialled the use of standardised colour cards (six colour shades) for both Indigenous and Caucasian groups to improve the reliability of vascularity assessment using a modified version of the VSS (Forbes-Duchart et al. 2007). A significant improvement in reliability was reported only for the Caucasian study group (refer to Appendix 6 for colour scales).

‘Good’ to ‘excellent’ inter-rater reliability ($k_w$ 0.72-0.85) for assessment of scar height was found in the study in this thesis using the plexiglass ruler as suggested by Baryza and Baryza (1995). This is greater than the inter-rater reliability reported in Baryza and Baryza (1995) ($k_w$ 0.56), Forbes-Duchart and others (2007) ($k_w$ 0.54-0.58) and Nedelec and others (2000) ($k_w$ 0.07-0.27). The use of calibrated silicone strips may be a useful alternative for measuring scar height to improve reliability, especially in anatomical locations where the plexiglass tool is difficult to manipulate (e.g. the axilla). Calibrated silicone strips (available in 1 mm, 2 mm and 5 mm thickness) are a new scar height measurement tool developed by Sian Falder and others at the Liverpool Alder Hey Children’s Hospital, UK (personal communication).

Pliability assessment using the mVSS demonstrated ‘excellent’ inter-rater reliability ($k_w$ 0.77-0.86) in this study. This is greater than the inter-rater reliability for pliability reported by Baryza and Baryza (1995) ($k_w$ 0.71), Forbes-Duchart and others (2007) ($k_w$ 0.12-0.38) and Nedelec and others (2000) ($k_w$ 0.19-0.22). While the mVSS descriptors used to classify scar pliability (e.g.
supple, yielding, firm, banding (‘rope-like’) and contracture) appeared to be interpreted consistently by raters in the study in this thesis, the very low level of inter-rater reliability in some of the other studies show that this is not always the case.

5.1.1.3 Conclusions

The reliability of the measurements obtained using the mVSS-TBSA method ranged from ‘marginal’ to ‘excellent’ and depended on the parameter being measured and the severity of the scar area being assessed. These results reflect the subjective nature of the method and point to particular challenges in measuring pigmentation and vascularity reliably. The mVSS-TBSA does demonstrate practical utility for assessing scars for both clinical and research purposes; however, there is potential to misclassify scar outcome. Protocol driven scar assessment to improve the reliability of the mVSS-TBSA method is recommended and as well as consideration of the use of objective scar assessment devices.

5.1.2 Reliability of the DermaLab Combo® (Study II – Chapter 3)

5.1.2.1 Summary of study

This study assessed the inter-rater and test-retest reliability of the DermaLab Combo® pigmentation, vascularity, and pliability and thickness measurements in burn scars. Three raters performed scar assessments on thirty patients with burn scars and inter-rater and test-retest reliability were assessed. Refer to Table 5-2 for summary of results.

Table 5-2: Key results of reliability testing of the DermaLab Combo®

- Pliability – ‘excellent’ test-retest reliability
- Thickness – ‘excellent’ inter-rater and test-retest reliability
- Pigmentation – ‘excellent’ inter-rater and test-retest reliability
- Vascularity – ‘good to excellent’ inter-rater reliability and ‘marginal’ test-retest reliability

5.1.2.2 Measurement issues

The DermaLab Combo® colour probe for pigmentation and erythema uses the measuring principle of spectrophotometry, reported to be a reliable and objective method for skin colour assessment (Stamatas et al. 2004; Takiwaki & Serup 1994). The measurement of scar pigmentation using the DermaLab Combo® demonstrated ‘excellent’ inter-rater (ICC 0.94-0.98), similar to the inter-rater reliability reported by Nedelec and others (2008) using a Mexameter (Khazaka electronics) (ICC 0.95-0.98) and Draaijers and others (2004b) using the
DermaSpectrometer (Cortex Technologies, Denmark) (ICC 0.94) and greater than the inter-rater reliability reported by Van der Kerckhove and others (2005) a using the Chromamater CR-300 (Minolta) (ICC 0.73-0.89). The test-retest reliability for pigmentation using the DermaLab Combo® was ‘excellent’ (ICC 0.83-0.87) as well; however, Nedelec and others (2008), Draaijers and others (2004b) and Van der Kerckhove and others (2005) did not test test-retest reliability.

In the case of erythema (vascularity) assessment using the DermaLab Combo®, inter-rater reliability ranged from ‘good’ to ‘excellent’ (ICC 0.66-0.84) similar to values reported by Draaijers and others (2004b) using the DermaSpectrometer (Cortex Technologies, Denmark) (ICC 0.72) and Van der Kerckhove and others (2005) a using the Chromamater CR-300 (Minolta) (ICC 0.75): The inter-rater reliability of the Mexameter (Khazaka electronics) was ‘excellent’ (ICC 0.87-0.97) (Nedelec et al. 2008). Test-retest reliability of erythema assessment using the DermaLab Combo® in this thesis was only ‘marginal’ (ICC 0.29-0.39). The other published literature on devices for the assessment of erythema (vascularity) do not report on test-retest reliability. Narrow-band reflectance spectrophotometry can detect very small changes in erythema influenced by minor movement, palpation or manipulation (Latreille et al. 2007; Reed et al. 1961). This sensitivity may have been responsible for the lower than expected reliability for the vascularity measurements using the DermaLab Combo® device and other spectrophotometric devices. It is possible that the DermaLab Combo® erythema measurements obtained after another measurement intervention (e.g. after suction measurement for pliability), or over time (e.g. test-retest reliability testing), may have been confounded by vascular responses independent of the scar. As another example of this, the physical pressure of the DermaLab Combo® probe against the skin during erythema measurement may elicit a blanching effect with drainage of the superficial capillary network, producing a temporary false pale scar colour (Hallam et al. 2013).

Suction is the basic measuring principle of the DermaLab Combo® elasticity probe and ‘excellent’ test-retest reliability was demonstrated. The suction mechanism used by the DermaLab Combo® distorts the skin and affects subsequent measurements of elasticity at the same location. The manufacturer’s recommendation of a washout period of one hour between measurements on the same location precluded the assessment of the inter-rater reliability of the elasticity probe in this study. However, Anthonissen and others (2012) reported excellent inter-rater reliability of the DermaLab® in measuring pliability, a device that uses the same principle as the DermaLab Combo® (Cortex Technologies, Denmark).

Some problems were encountered with the use of the DermaLab Combo® elasticity probe. Firstly, measurements of thick and rigid scars were not possible; this issue was also reported by
Anthonissen and others (2012). Secondly, difficulties were observed in obtaining successful pliability measurements on scars located on the legs; this may have been related to the greater thickness of the epidermis of the leg skin. The probe was also ineffective in obtaining an adequate grip on the normal skin site when the subjects had a rough growth of body hair. Grip on the normal skin may also have been compromised in the presence of lesser amounts of hair and a measurement bias introduced when calculating differences between scar and normal skin measurements.

The thickness of the scar (i.e. within and above the baseline of the skin) was assessed using the DermaLab Combo® ultrasound probe. While the manufacturer’s specification of the DermaLab Combo® indicated a 3.4 mm penetration capacity for measurement of thickness, a ceiling of 2.5 mm penetration capacity was observed in this study, with the subsequent inability to measure burn scars thicker than 2.5 mm. Cortex Technologies have indicated on their website (www.cortextechnologies.dk) that ultrasound probes with greater penetration capacity will be available in the near future which may increase the utility of the DermaLab Combo® for measuring burn scars.

5.1.2.3 Conclusion

Further work needs to be done to establish a protocol to improve the reliability of scar vascularity measurements by the DermaLab Combo®, focusing on what needs to be done to avoid confounding influences on vascularity. Based on the project execution and results, the order in which scar components are measured by the DermaLab Combo® should be stipulated in future protocols: erythema and pigmentation to be measured first, followed by thickness, with scar pliability measured last. Skin should be shaved if possible prior to the application of the pliability probe. Successful measurements may not be able to be obtained by the DermaLab Combo® for scars on the legs and scars that are thick for pliability and height, respectively.
5.1.3 Exploration of the DermaLab Combo® pigmentation and vascularity measurements in relation to the mVSS (Study III – Chapter 4)

5.1.3.1 Summary of study

The primary objective of this study was to explore the interpretation of the DermaLab Combo® measurements of scar pigmentation and vascularity and the data harmonisation with respective categories of the mVSS. Indices for scar pigmentation (melanin index percent, MI%) and scar vascularity (erythema index percent, EI%) were generated to represent a percent change relative to the patient’s matched normal skin using the DermaLab Combo® measurements of melanin and erythema. Results of the study are presented in Table 5-3.

Table 5-3: Interpretation of DermaLab Combo® pigmentation and vascularity measurements by mVSS classification

- **Pigmentation**
  - MI% values (percentage change relative to normal skin) generated from the DermaLab Combo® melanin index of scar and normal skin could be interpreted according to the mVSS pigmentation classification better than a ‘delta’ score (arithmetic difference between melanin index of scar and normal skin)
  - Classification was further improved by using both the generated MI% and EI% values

- **Vascularity**
  - EI% values (percentage change relative to normal skin) generated from the DermaLab Combo® erythema index of scar and normal skin could not be interpreted according to the mVSS vascularity classification

5.1.3.2 Methodological issues

The major issue and limitation of this study was the use of a subjective reference standard (mVSS) to which the DermaLab Combo® pigmentation and vascularity data were compared. The results identified disagreement between the two expert raters using the mVSS and this subsequently contributed to the limited examination of the validity of the DermaLab Combo® in measuring pigmentation and vascularity. Future studies of validation and data harmonisation of the DermaLab Combo® measurements of scar pigmentation and vascularity should use a superior reference standard to that used in this study that may include standardised colour cards, or consideration to histology of the skin (biopsy) as a ‘gold standard’ where possible.
The effective study sample size was reduced by approximately 20% as a result of the lack of agreement between the scar component classifications by the two expert raters. Such disagreement and sample size impacts have also been reported by Fadzil and others (2009) in testing an objective measurement (Konica Minolta Chromameter CR-400) with the subjective Psoriasis Area and Severity Index (PASI) erythema score in psoriasis. Where future studies involve multiple subjective expert opinions the sample size should be inflated to accommodate this. Consideration should also be given to designing studies that incorporate greater homogeneity of skin type (i.e. stratification by Fitzpatrick Skin Types) and also by scar maturity.

5.1.3.3 Conclusions

Quantifying percentage changes in melanin and erythema relative to matched normal skin was a way to improve the understanding of the DermaLab Combo® pigmentation measurements. The DermaLab Combo®-derived MI% values were able to be classified into pigmentation categories of the mVSS (mixed pigmentation category excluded), and the concordance of pigmentation classification was further improved by taking into account the EI% values. The DermaLab Combo® is a multi-component device with acceptable reliability and potential for improvement of reliability with protocol-driven measurements. However, while the measurement provides continuous numerical data that may prove useful for identifying change over time in scar pigmentation and vascularity, further work is required to understand the DermaLab Combo® measurements and to optimise the interpretation of these data.

5.2 Future testing of the DermaLab Combo® – considerations

5.2.1 Standardised protocols

To achieve robust scar measurements with improved reliability, protocols that minimise measurement bias and confounding and optimise the use of the DermaLab Combo® are required. Firstly, protocols should stipulate the order in which the burn scar component assessments should be made, and secondly, protocols should provide standardised assessment techniques for the use of the DermaLab Combo® to measure each scar component.
5.2.2 Improved experimental design

5.2.2.1 Homogeneity of study subjects

The results of Chapter 4 emphasised the importance of using more homogenous samples of study subjects (e.g. skin type, time since injury) when designing studies to validate the DermaLab Combo® for measuring burns scar pigmentation and vascularity, particularly in the absence of a ‘gold standard’. A homogeneous sample would reduce the measurement variability and standard errors of the estimates (Chapter 4).

The current project stratified the study subjects by the subjective mVSS classification of pigmentation and vascularity categories (Chapter 4). However, it is recommended that future validation studies using the DermaLab Combo® consider stratification of study subjects using several variables. This project identified that the inter-rater reliability of the mVSS depended on the severity of the scar (Chapter 2) and as such stratifying subjects by time since burn injury will allow greater understanding of the change in the DermaLab Combo® measurements in relation to scar progression. Stratification of study subjects by Fitzpatrick Skin Type may also provide an insight into the subsequent differences and impacts of skin type that may prove helpful in understanding and interpreting the DermaLab Combo® measurements.

5.2.2.2 Sample size

The sample size for the Study I (Chapter 2) and Study II (Chapters 3) were based on published recommendations for clinical reliability investigations (Whitley & Ball 2002; Walter et al. 1998) and for the final study (Chapter 4), on the sample size that would be required to undertake ROC analysis, if feasible (Microsoft MedCalc version 12.7.0 2013). Study III (Chapter 4) used the subjective mVSS as the comparison measure and the sample size should have been increased at the outset to accommodate the potential misclassification rates observed. However, the mVSS scores of two highly experienced clinical burn scar raters (with over 10 years of experience in the use of the mVSS) were used as the reference standard and the high mVSS misclassification rates observed were not anticipated. Further research should consider estimates of sample sizes with respect to multiple experts using a subjective reference standard, stratification by skin type and/or time since burn injury to support determination of reliability estimations with greater statistical confidence (Streiner & Norman 2008).

5.2.3 Improved reference standards for exploration of validity

In the absence of a suitable ‘gold standard’ in burn scar assessment, further reliability and validation studies of the DermaLab Combo® should consider the following options as reference standards for skin colour. Firstly, the use of standardised colour reference cards similar to those
used to quantify colour in relation to erythema by Forbes-Duchart and others (2007) for their modified version of VSS or those used in the Psoriasis Area and Severity Index (PASI) (Harari et al. 2009), or possibly, skin colour charts used in the cosmetic industry (e.g. Skin Color Chart® [L’Oréal] (de Rigal et al. 2007)); and, secondly, scar histology (biopsy). Refer to Appendix 6 for standardised colour reference charts.

5.3 Evidence generated by the project to guide scar assessment

5.3.1 Protocol-driven scar assessment

Results presented in Chapters 2, 3 and 4 highlight the need for protocol driven scar assessment when using either the subjective scar assessment scales or objective devices. Protocol-directed assessment will reduce systematic variation or bias in measurements and provide a safeguard for quality data collection for both clinical burn scar management and research (Hennekens & Buring 1987; McDowell & Newell 1996). Availability of high quality data will ultimately lead to better patient outcomes (Newman et al. 1998). Scar assessment in clinical burn care and research may often involve multiple raters with potential for measurement bias. Protocol driven scar assessment will improve the reliability (intra-rater, inter-rater and test-retest reliability) of the Scar assessment measurements with the best evidence driving clinical scar management.

Standardised protocols are essential for ongoing quality assurance training of experienced raters as well as for the upskilling and training of new scar assessors. Certification in burn scar assessment may also be sought, similar to that adopted by the international dermatological community for the training and certification of raters in the use of the PASI scale in psoriasis (PASI Training 2008).

5.3.2 Individual burn scar components or total global score?

Scar assessment parameters (pigmentation, vascularity, pliability and height) provide the clinically observable and measureable effects of the altered stages of wound healing that occur during the scar progression (Satish & Kathju 2010). Each parameter provides information on a distinct physiological process such as the altered pigmentation pattern, vascular response and fibrosis and represents information on a single dimension. The current scar assessment scales (e.g. VSS, mVSSs and POSAS) combine these single parameter scores to achieve a total score to represent the scar at the time of the assessment, with a higher score representing a worse scar outcome.
Health measurement instruments, including scar assessment tools, are generally designed for clinical use and to be sensitive to change following treatments (McDowell & Newell 1996). There are two contrasting methodological approaches that may be exercised when summarising data collected by instruments. Firstly, scores may be presented separately to represent the various dimensions of the health assessment (e.g. separate scores for scar pigmentation, vascularity, pliability and height) to provide a health profile (or scar profile). Secondly, scores of the individual dimensions or indicators of the aspect of health (e.g. scar) are aggregated into a single overall score or health index (e.g. total mVSS score for burn scar).

The single health index approach acknowledges that combining information from different health dimensions is problematic; however, it is argued that the trade-offs between information of the dimensions is often necessary in the decision making process of preferred treatment choices (McDowell & Newell 1996). Van der Wal and others (2012a) were the first to apply modelling (Rasch) to the Patient and Observer Scar Assessment Scale (POSAS) scar assessment tool. The authors concluded that the items/categories included in the POSAS, other than surface area, were uni-dimensional allowing for summation of scores into a meaningful total score. A search of the current literature did not identify such analyses that examined the VSS or its modifications.

The health profile approach is more commonly used by clinicians and its proponents support that health measurements, that would include scar assessment, are multidimensional and therefore the dimensional specific data should be presented separately (McDowell & Newell 1996). Tyack and others (2012) discuss the lack of evidence to support combining the four scar parameters into a total score in burn scar assessment. The authors argue that the total score does not adequately represent the sum of the scores for the individual parameters within the burn scar assessment scale. Most currently used scar assessment scales (Chapter 1) demonstrate low internal consistency and provide further support that the use of a total score is problematic. In relation to the mVSS, the mVSS pigmentation parameter is a nominal scale (normal pigmentation, hypopigmentation, mixed pigmentation and hyperpigmentation) and vascularity, pliability and height are represented using ordinal scales; combination of such scores to form a total mVSS score is not appropriate.

In summary, the use of total scores in scar assessment (e.g. mVSS) may need to be reconsidered. Data on individual scar parameters will provide useful and specific clinical information on the underlying physiological processes that will guide burn scar management and treatment interventions to optimise patient outcomes.
5.3.3 Measurements of burn components – ‘best’ test

Evidence has been generated from this project in relation to the reliability of the dimensions of both the subjective mVSS scale and the objective DermaLab Combo® multi-component device in measuring burn scar parameters. The results support that burn scar assessments would be enhanced by using the most reliable measures of the subjective mVSS scale and the objective DermaLab Combo® device.

Scleroderma is a dermatological condition that also uses scar assessment to guide clinical care. Liebling and others (2011) have recommended the use of both objective and subjective methods of scar assessments of localised scleroderma skin lesions. The widely used clinical assessment method – the modified Rodnan Skin Score (mRSS) is used alongside the objective ultrasound imaging for the early detection of localised scleroderma skin lesions (Liebling et al. 2011). Ultrasound imaging has the capacity to detect localised scleroderma lesions in both superficial and deep soft tissues early, before clinical signs of the disease are apparent. This is an advantage over the subjective mRSS method which can only detect scleroderma activity in deeper tissue in the presence of clinical symptoms (Liebling et al. 2011). Adoption of such an assessment approach in burn scar monitoring would be useful.

5.3.3.1 Assessment of burn scar components – evidence-based reliability

Until all burn scar parameters can be assessed by reliable and valid objective measures that may include an objective multi-component device, there is a need to use the most reliable components of the subjective and objective scar assessment measures available to inform burn care and research. A summary of the inter-rater reliability for burn scar parameters for the objective DermaLab Combo® multi-component device and the subjective mVSS scale are presented in Table 5.4.

<table>
<thead>
<tr>
<th>Scar Component</th>
<th>DermaLab Combo®</th>
<th>mVSS method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC</td>
<td>Weighted Kappa (k_w)</td>
</tr>
<tr>
<td>Pigmentation</td>
<td>0.94-0.98</td>
<td>0.02-0.57</td>
</tr>
<tr>
<td>(Melanin)</td>
<td>(‘excellent’)</td>
<td>(‘marginal to good’)</td>
</tr>
<tr>
<td>Vascularity</td>
<td>0.68-0.85</td>
<td>0.44-0.76</td>
</tr>
<tr>
<td>(Erythema)</td>
<td>(‘good to excellent’)</td>
<td>(‘good to excellent’)</td>
</tr>
<tr>
<td>Height†</td>
<td>0.93-0.99</td>
<td>0.72-0.85</td>
</tr>
<tr>
<td>(Thickness*)</td>
<td>(‘excellent’)</td>
<td>(‘good to excellent’)</td>
</tr>
<tr>
<td>Pliability</td>
<td>Not assessed</td>
<td>0.77-0.86</td>
</tr>
<tr>
<td>(Elasticity)</td>
<td></td>
<td>(‘excellent’)</td>
</tr>
</tbody>
</table>

† mVSS
*DermaLab Combo®
On the basis of the project results the following recommendations for measurements of the individual burn scar parameter are presented.

5.3.3.1.1 Pliability

Study results support the recommendation that pliability of burns scars be assessed using the mVSS. This project demonstrated ‘excellent’ inter-rater reliability of pliability assessment using the mVSS. The DermaLab Combo® pliability measurement had excellent test-retest reliability; assessment of inter-rater reliability was not possible during this project. However, Anthonissen and others (2012) have demonstrated ‘excellent’ inter-rater reliability for pliability measures using the DermaLab®, a device that uses the same technology as the DermaLab Combo® (refer to Chapter 1, Table 4). However, until a full understanding of the DermaLab Combo® pliability measures is achieved, the mVSS can be used with confidence to measure the pliability of burn scars.

5.3.3.1.2 Height/thickness

For the assessment of the height of burns scars it is recommended that both the mVSS and the DermaLab Combo® be used. The mVSS height measurement and the thickness measurement of the DermaLab Combo® are different and cannot be readily compared. The mVSS measures the height of a scar (section of scar above the baseline of the skin), while the DermaLab Combo® measures thickness of the scar which is within the skin and above the skin giving a more complete measure. The DermaLab Combo® measured thickness with ‘excellent’ reliability (both inter-rater and test-retest reliability) and the thickness measurement is more clinically informative than the height measurement of the mVSS. However, until a DermaLab Combo® ultrasound probe with a greater penetration capacity is available, the mVSS and the DermaLab Combo® both should be used for assessing the height/thickness of burn scars.

5.3.3.1.3 Pigmentation

Based on project results, it is recommended that the scar pigmentation assessment be conducted using both the mVSS and the DermaLab Combo®. The DermaLab Combo® pigmentation measurements demonstrated ‘excellent’ reliability (inter-rater and test-retest reliability); however, further work needs to be undertaken to facilitate the interpretation of these measurements.

5.3.3.1.4 Vascularity

Vascularity measurement using the mVSS and the DermaLab Combo® were similar (‘good’ to ‘excellent’ inter-rater reliability) however, the test re-test reliability of vascularity measurements using the DermaLab Combo® was ‘marginal’ to ‘good’. Until further testing and protocol development of the DermaLab Combo® vascularity measurements are undertaken, and
measurements are better understood, the mVSS is the most suitable method to assess burn scar vascularity.

5.3.4 Evidence-based hybrid protocol – using best of both worlds

Evidence-based assessment is accepted in clinical practice alongside effectiveness, safety and quality of care. Most fields of medicine now adopt knowledge gained from clinical trials to inform clinical decision making (Newman et al. 1998). Despite this trend, there are many clinical fields, including clinical scar assessment, where there is a lack of evidence to inform clinical practice (Fearmonti et al. 2010). This project has generated evidence in relation to both the subjective mVSS scale and the objective DermaLab Combo® multicomponent device in measuring burn scar parameters. Despite the requirement of further validation research of the DermaLab Combo® in measuring burn scar parameters; this project has generated valuable initial data to support the utility of the DermaLab Combo® in measuring some scar parameters. Based on this evidence a hybrid scar assessment protocol is proposed that uses scar parameters measured by the mVSS and/or the DermaLab Combo®, to maximise the reliability of scar assessment. Refer to Appendix 7 for the proposed evidence-based ‘hybrid’ burn scar assessment protocol.

5.4 Project strengths and limitations

5.4.1 Strengths

Burn scar outcome measures are important in the evaluation of patient functional and aesthetic outcomes and also are significant for research purposes. This is the first time an objective multicomponent device with capacity to measure four burn scar parameters (pigmentation, vascularity, pliability and thickness) has been identified and the reliability tested in clinical burn scar assessment. This project generated evidence that demonstrated that the DermaLab Combo® measured the scar parameters pigmentation, vascularity and thickness more reliably than the mVSS. However, despite the reliability of the DermaLab Combo® measurements, interpretation of the pigmentation, and pliability and vascularity data in the clinical context of burn scars needs further testing before the DermaLab Combo® can be included in routine clinical burn scar assessment.

Study III generated initial evidence to support the interpretation of the derived DermaLab Combo® pigmentation MI% values analogous to the clinical interpretation of the mVSS. Correlation analyses showed positive associations of moderate strength between the mVSS
clinical classification of pigmentation and vascularity with the respective DermaLab Combo®
derived MI% and EI% values were also found. These results supported that the mVSS scores
and DermaLab Combo® derived MI% and EI% values were assessing the same underlying
constructs. That is, as mVSS pigmentation classification (mixed pigmentation excluded)
increased from hypo- to normal- to hyperpigmentation, the DermaLab Combo® MI% values
increased. Likewise, as the mVSS erythema classification increased from normal to pink to red
through to purple, the DermaLab Combo® EI% values also increased. This provides a valuable
research platform from which further investigations can be conducted.

5.4.2 Limitations

Limitations related to measurements of burn scar parameters for both the mVSS and the
DermaLab Combo® and issues related to heterogeneity and sample size have been discussed
within each respective study presented and in this final chapter. However, the major limitation
of this project was the lack of available ‘gold standard’ in scar assessment (Fearmonti et al.
2010; Idriss & Maibach 2009). This is not an uncommon phenomenon in the health domain, and
the literature suggests that in the absence of a gold standard for validation of health instruments,
expert opinion is a reasonable proxy (Katz et al. 2000; Sullivan Pepe 2003; Fletcher et al. 1996).
For this project the current scar assessment method used at the RPH, a modification of the
Vancouver Scar Scale (mVSS) was used as the reference standard for testing the DermaLab
Combo®. Prior to the testing of the DermaLab Combo® in measuring burn scar parameters
(pigmentation, vascularity, pliability and height/thickness), the inter-rater reliability of the
mVSS (RPH version) was established (Chapter 2).

The mVSS is a subjective assessment and results in Chapter 4 emphasised the disagreement
between the two experienced clinical assessors using the mVSS. This subsequently limited
successful examination of the validity of the DermaLab Combo® in measuring pigmentation
and vascularity of burn scars. The use of other reference standards such as standardised colour
cards as discussed previously in Section 5.2.3 may prove to be useful. Consideration should also
be given to the histology of the skin (via biopsy) as a ‘gold standard’ where possible rather than
a reliance on subjective clinical opinion to validate objective devices.
5.5 Conclusions and recommendations

This project generated evidence to demonstrate that the DermaLab Combo® measured scar parameters pigmentation, vascularity and thickness more reliably than the mVSS. Initial evidence was also generated to support the interpretation of the MI % values (derived from the DermaLab Combo® pigmentation measurements) according to the mVSS classification. However, further work is required before the DermaLab Combo® measurements of pigmentation, vascularity and pliability can be translated successfully into clinical burn scar assessment. An important dimension of this research project is that it makes a valuable contribution to the body of knowledge associated with burns scar outcomes, with a particular emphasis on objective scar assessment.

Worldwide, most burn care facilities rely on subjective scar assessment methods to monitor scar progression. Subjective scales inherently yield burn scar data that are prone to measurement issues and lower reliability. However, these data are used to guide clinical decision making with regard to treatment and management of post burn scars; this may not be the best way forward. While the inter-rater reliability of the DermaLab Combo® measurements were superior to most mVSS scar measurements (pigmentation, vascularity and thickness), the clinical interpretation is limited.

Given the reliability of the numerical scar data generated by the DermaLab Combo® is greater for most scar components than the mVSS, it is prudent to use both the DermaLab Combo® and the mVSS concurrently until such a time that the DermaLab Combo® scar measurements are fully interpretable and translated. Therefore, informed by the results of this project, a hybrid method of scar assessment that incorporates reliable measures of both subjective and objective scar assessment method and standardised measurement protocols is proposed. It is anticipated that such a hybrid burn scar assessment method will achieve burn scar data of higher quality to guide scar monitoring, clinical decision-making and treatment.

The following recommendations are based on the findings of this research:

**Recommendation one**

Evidenced-based scar assessment that incorporates the most reliable subjective and objective measures of the burn scar parameters in a hybrid method should be used until all burn scar parameters can be measured with reliable, valid and objective device(s).

**Recommendation two**

The use of individual scar parameter scores to inform specific clinical information on the underlying physiological process of the scar should be used rather than total scores in scar
assessment (i.e. mVSS and other subjective scales) to guide scar monitoring and treatment to optimise patient outcomes.

**Recommendation three**

Protocol driven scar assessment be administered when using both subjective and objective scar assessment methods. Standardised protocols must be established within a burn care facility for the scar assessment methods used. These standardised protocols should be utilised for training of new scar assessors as well supporting quality assurance programs of experienced scar assessors.

**Recommendation four**

Consideration to be given to establishing global standardised burn scar assessment protocol(s) for currently used and more reliable burn scar assessment methods. This will ensure the compatibility of internationally collected burn scar data in the burn clinical management and research.

**Recommendation five**

Further testing of the DermaLab Combo® in burn scar assessment using robust experimental design is essential to facilitate translation of this multi-component device into the clinical assessment of burn injury and subsequent burn care and management.
References

Aarabi, S, Longaker, MT & Gurtner, GC 2007, 'Hypertrophic scar formation following burns and trauma: new approaches to treatment', *PLOS Med*, vol. 4, no. 9, pp. 1464-1470.


Balakrishnama, S & Ganapathiraju, A 1999, *Linear discriminant analysis-a brief tutorial*, Institute of Signal and Information Processing, Mississippi State University.


Clarys, P, Alewaeters, K, Lambrecht, R & Barel, AO 2000, 'Skin color measurements: comparison between three instruments: the Chromameter(R), the DermaSpectrometer(R) and the Mexameter(R)', Skin Res Technol, vol. 6, no. 4, pp. 230-238.


Cubison, TCS, Pape, SA & Parkhouse, N 2006, 'Evidence for the link between healing time and the development of hypertrophic scars (HTS) in paediatric burns due to scald', Burns, vol. 32, no. 8, pp. 992-999.


Hennekens, CH & Buring, JE 1987, Epidemiology in medicine, Little, Brown & Co.


Li-Tsang, CWP, Lau, JCM & Liu, SKY 2003, 'Validation of an objective scar pigmentation measurement by using a spectrocolorimeter', Burns, vol. 29, pp. 779-784.


Miller, M & Nanchahal, J 2005, 'Advances in the modulation of cutaneous wound healing and scarring', *Biodrugs*, vol. 19, no. 6, pp. 363-381.

Murray, BC & Wickett, RR 1997, 'Correlations between Dermal Torque Meter(R), Cutometer(R) and Dermal Phase Meter(R) measurements of human skin', *Skin Res Technol*, vol. 3, no. 2, pp. 101-106.


APPENDICES
Appendix 1: Supplementary Figure 1
Data collection sheet for the mVSS-TBSA method used at RPH (Chapter 2)
Appendix 2: Bland Altman Plots
Bland Altman plots for the continuous variables for all rater combinations (Chapter 2).

Please note for all the graphs:

- The dotted line denotes the mean difference between raters
- The grey shading denotes area defined by upper and lower limits of agreement.

‘Best’ scar area total mVSS score (see Table 2-3)
‘Worst’ scar area total mVSS score (see Table 2-3)
%TBSA allocation (see Table 2-4)
Less than 5 mVSS severity allocation (see Table 2-4)
5-10 mVSS severity allocation (see Table 2-4)
Appendix 3: Scar Assessment Form
**ROYAL PERTH HOSPITAL BURNS UNIT**

**DERMALAB COMBO - SCAR ASSESSMENT**

<table>
<thead>
<tr>
<th>Colour</th>
<th>Elasticity (PD only)</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best</td>
<td>Best</td>
</tr>
<tr>
<td></td>
<td>Melanin</td>
<td>Erythema</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worst</td>
<td>Worst</td>
<td>Worst</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>Control</td>
<td>Control</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Scar assessment for used for data collection (Chapter 3)
Appendix 4: Data Stratification Sheet
Data stratification sheet used in the third and final clinical trial (Chapter 4)

The DermaLab Combo Study Patient stratification sheets

Please affix a patient sticker appropriate to the category for each patient

### Pigmentation

<table>
<thead>
<tr>
<th>Normal (0)</th>
<th>Hypopigmentation (1)</th>
<th>Mixed pigmentation (2)</th>
<th>Hyperpigmentation (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Vascularity

<table>
<thead>
<tr>
<th>Normal (0)</th>
<th>Pink (1)</th>
<th>Red (2)</th>
<th>Purple (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 5: Data Collection Sheet
Data collection sheet used in the third and final clinical trial (Chapter 4)

## DermaLab Combo Study - Data Collection sheet

Affix patient sticker here

Date ______________________

<table>
<thead>
<tr>
<th>DermaLab Combo assessment</th>
<th>DermaLab Combo assessment</th>
<th>mvSS assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start time _____</td>
<td>End time _____</td>
<td>Start time _____</td>
</tr>
<tr>
<td>Scar</td>
<td>Melanin index</td>
<td>Normal skin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Average</td>
<td>Average</td>
<td>3</td>
</tr>
<tr>
<td>Scar</td>
<td>Erythema index</td>
<td>Normal skin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascularity</td>
<td>Erythema index</td>
<td>Vascularity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Average</td>
<td>Average</td>
<td>3</td>
</tr>
</tbody>
</table>

145
Appendix 6: Colour Charts
Colour charts (Chapter 5)

i) Forbes-Duschart and others (2007) used standardised colour charts in their modified version of the VSS for assessing erythema

Photograph of the modified VSS. Left: Caucasian scale and Middle: Aboriginal scale [Adopted from Forbes-Duschart et al. (2007)].

ii) Psoriasis Area and Severity Index (PASI) uses standardised colour cards for assessing erythema (Fadzil et al. 2009)

PASI colours of erythema. Colours representing each PASI erythema score and skin colour. (a) light pink (b) pink (c) red (d) dark red and (e) purple [Adopted from Fadzil et al. (2009)].
iii) Skin Color Chart ® (www.lorealusa.com) – skin colour charts used in the cosmetic industry as suggested by (de Rigal et al. 2007).

Photograph of the Skin Color Chart® displaying 66 skin tones developed by L’Oreal [Adopted from www.lorealusa.com].
Appendix 7: Hybrid Scar Assessment Protocol
Proposed evidence-based hybrid protocol

Prior to scar assessment

- Remove all pressure garments or bandages at least 10 minutes prior to assessment
- Patients should be in a comfortable position with the body part with the scar in a non-dependent position (seated or lying down) for a minimum of 15 minutes prior to the assessment. The position of the patient needs to be noted on the assessment sheet
- Select and mark with a semi-permanent marker a scar and adjacent normal skin area both 3 cm x 3 cm in size for assessment and photograph the marked areas
- Ensure all scar assessments are carried out in the same assessment room, with controlled environmental conditions (temperature range between 22-25°C)
- Record measurements in the scar assessment sheet (see page 161)
- Observe a wash-out period of 20 minutes between each measurement

Scar Identification:

- Indicate scar outline and the 3 cm x 3 cm areas for assessment on the body map
- Estimate the scar TBSA% using the ‘Rule of 9’s’ and record on the assessment sheet

Assessment of scar parameters

The order of scar parameters when assessing scars:

1) Pigmentation – DermaLab Combo®
2) Vascularity – mVSS
3) Pigmentation – mVSS
4) Thickness – DermaLab Combo®
5) Height – mVSS
6) Pliability – mVSS
Step 1
Pigmentation assessment using the DermaLab Combo®

Press the pigmentation button on the main screen of the device (refer to the image below) to open the colour analysis screen.

Use the DermaLab Combo® colour probe (as per Figure 1) by pressing the measure button to make a measurement.

Make three measurements within the scar areas following the probe placement guidelines given in Figure 1.

Record the three individual scar scores for melanin (melanin 1, melanin 2 and melanin 3) and the device generated average score for melanin (melanin average) on the assessment sheet (see Figure 2 for example screen).

Repeat for normal skin areas.

Do not apply any pressure other than the weight of the probe on the scar/normal skin when measuring.
Figure 1: Probe placement guidelines within the 3 cm x 3 cm measurement area. The circles represent the areas where the probes will be applied for each measurement.

Figure 2: The DermaLab Combo® colour analysis screen
Step 2

Vascularity assessment using the mVSS

Assess the amount of ‘redness’ in the 3 cm x 3 cm scar area in comparison to the adjacent normal non-injured skin.

Score vascularity as follows:

0=normal - colour that closely resembles the colour over the rest of the body
1=pink
2=red
3=purple

Record the vascularity score on the assessment sheet.

Step 3

Pigmentation assessment using the mVSS

To assess pigmentation apply mild pressure with the clear plexiglass tool to the assessment area until the first instance of skin blanching is achieved and compare the ‘brownness’ of the assessment area with the normal adjacent non-injured skin.

Score pigmentation as follows:

0=normal
1=hypopigmentation
2=mixed pigmentation
3=hyperpigmentation

Record the pigmentation on the assessment sheet.
Step 4
Thickness assessment using the DermaLab Combo®

Press the ultrasound button on the main screen of the device to open the thickness analysis screen (refer to the image below).

Main screen

A brief description of the DermaLab Combo® ultrasound screen:
A – Active image window indicated by the green stripe

B – Image window, which can be made active by clicking inside the window

C – Measure button

D – Intensity score (not relevant to this measurement)

E – Line grid which is activated upon pressing the 'line on' button (middle of the screen)

F – Buttons that move the grid once the 'line on' button is activated

G – Dermal thickness measure

H – Gain up and down buttons.

On the ultrasound screen is open the active window for measurement will be indicated by a green line (A) (refer to the image below).

Use the active window for first measurement. Touch the next window (B) to activate and use window (B) for the second measurement.

Apply a drop of gel on to the assessment area and spread the gel evenly.

Do not apply any pressure on the skin other than the weight of the probe when using the ultrasound probe.

Place the ultrasound probe on the skin and make three measurements on each 3 cm x 3 cm measurement area guided by the probe placements in Figure 1.

Press the measure button to make a measurement. The new image will appear on the active window (window with the green line).

Press the gain button to get the best possible image and repeat measures to get the best image.
Calculate dermal thickness according to the manufacturer’s instructions (http://www.cortex.dk/skin-analysis-products/multi-parameter-skin-testing.html).

Record the thickness measurements on the assessment sheet.

Repeat for normal skin areas.

Do not apply any pressure other than the weight of the probe on the scar/normal skin when measuring.
**Step 5**

**Height assessment using the mVSS**

Palpate the scar area in comparison to the normal adjacent non-injured skin area.

Use the clear Plexiglass ruler to assist in determining the height of the scar. (or calibrated silicone strips if available)

Measure the highest point if scar is uneven.

Score height as follows:

0 = normal – flat

1 = >0 to 1 mm

2 = >1 to 2 mm

3 = >2 to 4 mm

4 = >4 mm

Record the height score on the assessment sheet.

**Step 6**

**Pliability assessment using the mVSS**

Palpate the scar area and compare to the adjacent normal non-injured skin.

The rater should not wear gloves.

Score pliability as follows:

0 = normal

1 = supple – flexible with minimal resistance

2 = yielding – giving way to pressure

3 = firm-inflexible, not easily moved, resistant to manual pressure

4 = banding – rope-like tissue that blanches with extension of scar

5 = contracture – permanent shortening of scar producing deformity or distortion.

Record the pliability score on the assessment sheet.
Scar assessment sheet

Date ____________________________

Affix patient sticker here

Mark the contour of the scar being assessed on the body map.

Use an X to indicate the actual assessment location (3 cm x 3 cm square area) of the scar in relation to the whole scar (use the distance from a landmark, i.e. axilla, elbow etc.) and indicate the landmark on the table below. When assessing multiple scars use X1, X2 etc.

Estimate the %TBSA scar area using the ‘Rule of 9s’ for each scar assessed. Indicate in Table below.

<table>
<thead>
<tr>
<th>Date</th>
<th>Scar number</th>
<th>Landmark on the body</th>
<th>% scar TBSA</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

161
Step 1: The DermaLab Combo® pigmentation assessment

<table>
<thead>
<tr>
<th>Scar</th>
<th>Melanin index</th>
<th>Melanin index</th>
<th>Melanin index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scar 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scar 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scar 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanin 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanin 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanin 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal skin</td>
<td>Melanin index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanin 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanin 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanin 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 2: Vascularity assessment using the mVSS

<table>
<thead>
<tr>
<th>Scar 1</th>
<th>Scar 2</th>
<th>Scar 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tick</td>
<td>Tick</td>
<td>Tick</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scar</th>
<th>Tick</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (0)</td>
<td></td>
</tr>
<tr>
<td>Pink (1)</td>
<td></td>
</tr>
<tr>
<td>Red (2)</td>
<td></td>
</tr>
<tr>
<td>Purple (3)</td>
<td></td>
</tr>
</tbody>
</table>

Step 3: Pigmentation assessment using the mVSS

<table>
<thead>
<tr>
<th>Scar 1</th>
<th>Scar 2</th>
<th>Scar 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tick</td>
<td>Tick</td>
<td>Tick</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scar</th>
<th>Tick</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (0)</td>
<td></td>
</tr>
<tr>
<td>Hypopigmentation (1)</td>
<td></td>
</tr>
<tr>
<td>Mixed pigmentation (2)</td>
<td></td>
</tr>
<tr>
<td>Hyperpigmentation (3)</td>
<td></td>
</tr>
</tbody>
</table>
Step 4: Thickness assessment using the DermaLab Combo ultrasound probe

<table>
<thead>
<tr>
<th>Thickness DermaLab Combo® assessment (ultrasound probe)</th>
<th>Scar 1</th>
<th>Scar 2</th>
<th>Scar 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal skin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 5: Height assessment using the mVSS

<table>
<thead>
<tr>
<th>Height assessment using the mVSS</th>
<th>Scar 1</th>
<th>Scar 2</th>
<th>Scar 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scar</td>
<td>Tick</td>
<td>Tick</td>
<td>Tick</td>
</tr>
<tr>
<td>Normal (0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 1 mm (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 to 2 mm (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 to 4 mm (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;4 mm (4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 6: Pliability assessment using the mVSS

<table>
<thead>
<tr>
<th>Pliability assessment using the mVSS</th>
<th>Scar 1</th>
<th>Scar 2</th>
<th>Scar 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scar</td>
<td>Tick</td>
<td>Tick</td>
<td>Tick</td>
</tr>
<tr>
<td>Normal (0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supple- flexible with minimal resistance (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yielding- giving way to pressure (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm- inflexible, not easily moved, resistant to manual pressure (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banding- rope-like tissue that blanches with extension of scar (4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contracture- permanent shortening of scar producing deformity or distortion (5)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix A: A modified Vancouver Scar Scale linked with TBSA (mVSS-TBSA): Inter-rater reliability of an innovative burn scar assessment method (published work)
A modified Vancouver Scar Scale linked with TBSA (mVSS-TBSA): Inter-rater reliability of an innovative burn scar assessment method

T.U. Gankande a,*, F.M. Wood a,b, D.W. Edgar a,b, J.M. Duke a, H.M. DeJong a,b, A.E. Henderson b, H.J. Wallace a

a Burns Injury Research Unit, School of Surgery, The University of Western Australia, Australia
b Burns Service of Western Australia, Royal Perth Hospital, Australia

A R T I C L E   I N F O

Article history:
Accepted 28 January 2013

Keywords:
Burns
Scar assessment
Inter-rater reliability

A B S T R A C T

Background: Current scar assessment methods do not capture variation in scar outcome across the burn scar surface area. A new method (mVSS-TBSA) using a modified Vancouver Scar Scale (mVSS) linked with %TBSA was devised and inter-rater reliability was assessed.

Method: Three raters performed scar assessments on thirty patients with burn scars using the mVSS-TBSA. Scoring on pigmentation, vascularity, pliability and height was undertaken for the 'best' and 'worst' areas of each scar. Raters allocated the total body surface area of the scar (%TBSA) to three mVSS categories (<5, 5–10, >10). Intra-class correlation coefficient (ICC) and weighted kappa statistic (k_w) were used to assess inter-rater reliability. The data were also analysed for clinically relevant misclassifications between pairs of raters.

Results: Total mVSS scores showed 'fair to good' agreement (ICC 0.65–0.73) in the 'best' area of the scar while there was 'excellent' agreement in the 'worst' scar area (ICC 0.85–0.88). The k_w of the individual mVSS components ranged from 0.44 to 0.84 and 0.02 to 0.86 for 'best' and 'worst' scar areas, respectively. Determination of scar %TBSA had 'excellent' reliability (ICC 0.91–0.96). Allocation of scar %TBSA to severity category <5 mVSS demonstrated 'good to excellent' reliability (ICC 0.63–0.80) and 'fair to good' reliability (ICC 0.42–0.74) for 5–10 mVSS category. However, misclassifications were observed for the total mVSS score in the 'worst' scar area and the allocation of scar %TBSA in the <5 mVSS category.

Conclusion: Inter-rater reliability of mVSS scores depends on the severity of the scar area being assessed. The mVSS-TBSA method of allocation of scar %TBSA to two broad mVSS categories, namely <5 and >5 mVSS, has 'good to excellent' reliability. The mVSS-TBSA has demonstrated utility for both clinical and research purposes; however, there is potential to misclassify scar outcome in some cases.

© 2013 Elsevier Ltd and ISBI. All rights reserved.

1. Introduction

Cutaneous healing in all but the most superficial wounds results in a scar. Excessive or 'hypertrophic' scar formation is more common in post burn injuries than other traumatic injuries to the skin [1–3] and one of the main objectives of treatment is to minimise scarring [1]. All interventions from the point of injury aim to improve functionality, aesthetics and psychological outcomes to give a better quality of life after burn injury [3–5].

* Corresponding author at: Burn Injury Research Unit, School of Surgery, M318, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia. Tel.: +61 0 8 6488 8772; fax: +61 0 8 6488 8580.
E-mail address: ugangkade@meddent.uwa.edu.au (T.U. Gankande).
0305-4179/$36.00 © 2013 Elsevier Ltd and ISBI. All rights reserved.
http://dx.doi.org/10.1016/j.burns.2013.01.014
Monitoring the progression of a scar using some form of scar assessment is essential to evaluate the impact of therapy and the need to modify treatment if necessary [6]. Scar assessments are also fundamental for measuring research outcomes, allowing the comparison of factors that may influence scar outcomes as well as the assessment of clinical interventions.

Scar assessment scales are subjective in nature and provide a qualitative measure of the scar by the person performing the assessment (clinician/observer/patient) and have low inter-rater reliability [7]. Despite the limitations associated with scar scales, they are still used widely in the clinical setting because they are cheap, time-efficient and do not require much training [8].

The Vancouver Scar Scale (VSS) is the first validated burn scar assessment scale and remains the most widely used scale within the clinical setting [7]. In order to address issues associated with the use of the VSS, such as subjectivity and low inter-rater reliability, many authors have designed modifications to the VSS over the last few years. At present there are at least eight modified VSS in the published literature [9].

Although the VSS and its subsequent modifications are a step towards providing a framework for outcome assessment, they only measure of a single area within the scar and do not capture the variation in pathology across the surface area of the scar [10]. We propose a different system of scar assessment based on the VSS Baryza modification (mVSS) [11]. The mVSS-TBSA is a new method of application of the mVSS that links the mVSS with the assessment of % total body surface area (%TBSA) of the scar based on the ‘Rule of 9s’ [12]. The proposed new application has been developed to provide both improved description of the ‘whole’ scar and more complete information to enable monitoring of scar progression over time. For example, if at one month post-injury a 10% TBSA scar contains an area of 2% TBSA with VSS score of 10, and at 6 months post-injury the area with VSS score of 10 has reduced to 0.5% TBSA, an improvement of 1.5% TBSA can be monitored using this mVSS-TBSA, even though the VSS score of the ‘worst’ scar area remains static at 10 (Fig. 1).

A study to assess the inter-rater reliability of this new mVSS-TBSA was undertaken at Burns Service of Western Australia, Royal Perth Hospital, Western Australia. Our expectation was that three raters, with a range of duration of experience in scar assessment, could measure scar outcome reliably using the new method.

2. Methods

2.1. Subjects

A group of 30 adult patients scheduled for scar assessments as part of their routine follow-up care post burn injury, were recruited. One index scar per patient (see Section 2.4) was assessed for a total of 30 scars. Subjects were all volunteers and met the following eligibility criteria: at least 18 years of age; presence of a healed immature burn scar less than 18 months post injury; scar area of at least 6 cm × 3 cm; able to provide informed written consent. Subjects who had other dermatological conditions in conjunction with their burn injury were excluded from the study. Written informed consent was obtained after providing subjects with a Patient Information Sheet and a verbal explanation. The study was approved by the Human Research Ethics Committee of Royal Perth Hospital (RPH) and The University of Western Australia.

2.2. Study design

The study was designed as a prospective correlation study to assess the inter-rater reliability of the mVSS and %TBSA allocation as part of the new mVSS-TBSA method. Three raters assessed 30 individual index scars on 30 patients, that is, one scar per patient. Within each individual index scar the ‘best’ scar area of 3 cm × 3 cm and the ‘worst’ scar area of 3 cm × 3 cm were assessed using the mVSS. Therefore, the raters assessed 60 scar sites in total. Raters then allocated the total body surface area of the scar (%TBSA) to three mVSS categories (<5, 5–10, >10). Each rater was blinded to the results.
of other raters. The person performing the data analysis was blinded to the identity of each rater.

2.3. Raters

All raters had conducted at least 50 assessments using the new mVSS-TBSA prior to the commencement of the study. The raters included a senior occupational therapist with over five years of experience of the method, a second senior occupational therapist with over eight months experience of the method, and a clinician researcher with six months experience of the method. For the remainder of the paper, we will refer to these raters as R1, R2 and R3.

2.4. Procedure

All pressure garments and bandages were removed at least 15 min before the start of the first assessment. In the event that a patient had more than one scar the anatomical body part with the highest %TBSA scar (healed) was selected as the index scar for assessment. The index scar (right arm, left arm, back, chest, right leg, left leg, head, neck and groin) was identified by the most experienced rater. A 3 cm × 3 cm scar area deemed to be the ‘best’ area of the scar (lowest mVSS score) and 3 cm × 3 cm area of the same scar deemed to be as the ‘worst’ area of the scar (highest mVSS score) were identified by each rater, and rated based on the mVSS (Baryza modification) [11]. This modification incorporates a Flexiglas tool for blanching the scar and measuring scar height. Each rater then used the ‘Rule of 9s’ to estimate the scar area (%TBSA) and allocated the % TBSA to three mVSS severity categories (<5 mVSS, 5–10 mVSS, >10 mVSS). The three independent raters performed the scar assessments on each subject on the same day. The raters were blinded to the results recorded by other raters.

2.5. Data collection

Data were collected over a period of 5 weeks. The results were recorded in the Assessment Form which comprises the numerical mVSS scale (Table 1), a body map and a chart for allocation of the % TBSA (Appendix I—Supplementary Fig. 1) [12]. All forms were filed separately until the data collection was complete.

2.6. Data analysis

Descriptive statistics including age, gender, scar % TBSA, anatomical location of the scar, and time since injury, were calculated. Separate inter-rater reliability assessments were made for the ‘best’ and ‘worst’ scar areas for each mVSS component and the total mVSS score. The inter-rater reliability of determining the % TBSA of the whole scar and the allocation of the scar % TBSA to three severity categories (<5 mVSS, 5–10 mVSS, >10 mVSS) was also assessed. Data analyses were performed using SPSS version XX (Chicago, Inc) and Stata V12 statistical software (StataCorp LP, College Station, TX).

The Bland Altman method was used for continuous variables (total mVSS score, scar %TBSA and the allocation of scar %TBSA to the three severity categories) and plots of the difference in scores versus average of scores, for each pair of raters (R1 vs. R2, R1 vs. R3, and R2 vs. R3) were generated with mean difference and 95% limits of agreement overlaid [13]. The 95% limits of agreement represent limits of how different the measurements of two observers could plausibly be for a subject and were calculated as the mean of the differences ±2 × standard deviation. The proportion of values falling within the 95% limits of agreement was reported for each rater combination. Plots were inspected for any tendency for the amount of variation to change with the magnitude of the measurement. If no such tendency was observed, inter-rater reliability was assessed using two-way random effects ANOVA was used to derive the intra-class correlation coefficients (ICC) with 95% confidence intervals for each pair of raters.

The ICC is a measure of agreement that records the average similarity or reproducibility of raters’ actual scores on the ratings being compared [14]. The ICC is also a measure of the proportion of total variation that is between subjects variation. Therefore, a limitation of the ICC is that if variation due to observers is large compared with the variation between subjects then ICC will be closer to zero [15]. Different guidelines exist for the interpretation of the ICC, however, one acceptable scale is that an ICC value of less than 0.40 indicates poor reproducibility or agreement, ICC values in the range of 0.40 to 0.75 indicate fair to good agreement, and an ICC value of greater than 0.75 demonstrates excellent agreement [16,17].

For the discrete variables of the mVSS components (pigmentation, vascularity, pliability, height), inter-rater reliability was assessed using the weighted kappa statistic (κw) with quadratic weights [18] for each pair of raters (R1 vs. R2, R2 vs. R3 and R1 vs. R3). The calculation of the kappa statistic is

<table>
<thead>
<tr>
<th>Table 1 - Numerical scale of mVSS used in the study.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pigmentation</td>
</tr>
<tr>
<td>0 = normal</td>
</tr>
<tr>
<td>1 = hypo-pigmentation</td>
</tr>
<tr>
<td>2 = mixed pigmentation</td>
</tr>
<tr>
<td>3 = hyper-pigmentation</td>
</tr>
<tr>
<td>Vascularity</td>
</tr>
<tr>
<td>0 = normal</td>
</tr>
<tr>
<td>1 = pink</td>
</tr>
<tr>
<td>2 = red</td>
</tr>
<tr>
<td>3 = purple</td>
</tr>
<tr>
<td>Pliability</td>
</tr>
<tr>
<td>0 = normal</td>
</tr>
<tr>
<td>1 = supple—flexible with minimal resistance</td>
</tr>
<tr>
<td>2 = yielding—giving way to pressure</td>
</tr>
<tr>
<td>3 = firm—inflexible, not easily moved,</td>
</tr>
<tr>
<td>resistant to manual pressure</td>
</tr>
<tr>
<td>4 = banding—rope-like tissue that blanches</td>
</tr>
<tr>
<td>with extension of scar</td>
</tr>
<tr>
<td>5 = contracture—permanent shortening of</td>
</tr>
<tr>
<td>scar producing deformity or distortion</td>
</tr>
<tr>
<td>Height</td>
</tr>
<tr>
<td>0 = normal—flat</td>
</tr>
<tr>
<td>1 = &gt;0 to 1 mm</td>
</tr>
<tr>
<td>2 = &gt;1 to 2 mm</td>
</tr>
<tr>
<td>3 = &gt;2 to 4 mm</td>
</tr>
<tr>
<td>4 = &gt;4 mm</td>
</tr>
<tr>
<td>Total 1</td>
</tr>
</tbody>
</table>
based on the difference between how much agreement is actually present (‘observed’ agreement) compared to how much agreement would be expected to be present by chance alone (‘expected’ agreement) [18]. In addition to $k_w$, for height assessments in the ‘best’ scar area, observed and expected agreement, as well as the observed proportions of agreement for height category = 0 and height category = 1, as defined by Cicchetti and Feinstein [19], were reported for rater combinations R1 vs. R2 and R2 vs. R3. The interpretation of $k_w$ values by Rosner [16] is that 0–0.4 indicates marginal agreement, 0.4–0.75 is indicative of good agreement while a score over 0.75 demonstrates excellent agreement.

Evaluation of inter-rater reliability requires interpretation in the clinical context to be meaningful [20]. Therefore, the data were analysed for clinically relevant misclassifications between pairs of raters. An a priori decision was made that a difference of 2 units in score for the vascularity, pliability and height components represented a clinically relevant misclassification (refer to Table 1). For the total mVSS score, misclassification based on a total score <5 and ≥5 mVSS was used. The allocation of scar %TBSA to mVSS severity categories (<5, 5–10, >10) was considered a misclassification for a difference of >1% scar TBSA between raters. The pigmentation component was not included in this analysis as it is a nominal scale.

3. Results

3.1. Subjects

Among the 30 subjects, 22 (73.3%) were male and the median age of study subjects was 30.5 years (inter-quartile range [IQR], 26.7–53.0 years). The median scar % TBSA was 1.5% (IQR 1.0–4.8%) and the median time since injury among subjects was 3.0 months (IQR 1.8–6.5 months). The locations of the assessed scars were: arm ($n = 11$), leg ($n = 12$), chest ($n = 3$), head and neck ($n = 2$), and back ($n = 2$).

3.2. mVSS scores

Fig. 2 summarises the total mVSS scores derived by all raters in the ‘best’ and ‘worst’ scar areas. There were only a few scores above mVSS 10, with the majority of mVSS scores being in the <5 and 5–10 mVSS categories. As expected, there were more scores in the 5–10 and >10 mVSS categories in the ‘worst’ scar area compared to the ‘best’ scar area.

3.3. Inter-rater reliability

Table 2 presents pair-wise weighted kappa statistics for each mVSS component in the ‘best’ and ‘worst’ area of the scar. Rater agreement with respect to the mVSS components was variable. Rater agreement on pigmentation assessment was ‘marginal’ in the ‘worst’ scar area ($k_w$ 0.30, 0.02, 0.33 for each rater pair) and ‘good’ in the ‘best’ scar area ($k_w$ 0.57, 0.45, 0.54 for each rater pair). Pliability assessment achieved ‘excellent’ agreement with $k_w$ scores of 0.83, 0.84 and 0.82 (‘best’ scar area) and 0.77, 0.86 and 0.84 (‘worst’ scar area), respectively. Height assessment showed ‘good to excellent’ agreement in the ‘worst’ scar area ($k_w$ 0.85, 0.72 and 0.74). All height measurements recorded in the ‘best’ scar area were either category 0 or 1. A high level of agreement was observed between all rater pairs. The $k_w$ scores were unable to be determined for pair-wise comparisons with R2 because R2 recorded all scores as 0. Rater pair (R1 vs. R3) with a $k_w$ score 0.76 demonstrated ‘excellent’ agreement. Vascularity assessment was indicative of ‘good to excellent’ agreement in ‘worst’ scar area ($k_w$ 0.64–0.76) and in the ‘best’ scar area ($k_w$ 0.44–0.71).

Bland Altman plots for the continuous variables (total mVSS score in ‘best’ and ‘worst’ areas, scar %TBSA and allocation of scar %TBSA in three severity categories: <5 mVSS, 5–10 mVSS, >10 mVSS) were inspected for systematic patterns. No tendency for variation to change with the magnitude of the measurement was observed (Appendix II). For all Bland Altman plots of difference in scores versus the mean of the score for each rater pair, almost all values fell within the limits of agreement.

ICCs and Bland Altman values (mean difference, upper and lower 95% limits of agreement, proportion within limits of agreement) for pairs of raters for total mVSS scores and scar %TBSA and mVSS category allocation are presented in Tables 3 and 4, respectively. The total mVSS score had ‘good’ rater agreement in the ‘best’ scar area (ICC 0.65–0.73) for all the rater combinations. In the ‘worst’ area of the scar, there was ‘excellent’ rater agreement (ICC 0.85–0.88) for all the rater combinations. The rater agreement in determining the %TBSA of the index scar was ‘excellent’ (ICC 0.90) for all rater

![Fig. 2 – Allocated mVSS scores by raters (R1, R2 and R3) in ‘best’ and ‘worst’ areas of the index scar.](image)
combinations. The agreement between pairs of raters for allocation of the scar %TBSA to mVSS categories (<5 mVSS, 5–10 mVSS, >10 mVSS) varied (Table 4). Results of the ICC supported ‘good’ agreement for R1 vs. R3 (ICC 0.63) in allocating %TBSA to the <5 mVSS category and ‘excellent’ agreement for R1 vs. R2 and R2 vs. R3 (ICC 0.80, 0.78). The allocation of scar %TBSA to the 5–10 mVSS category showed ‘fair to good’ agreement (ICC 0.42–0.74).

In Table 3, the Bland Altman mean differences for total mVSS scores varied from −0.50 to 0.57 and −1.0 to 0.37 for ‘best’ and ‘worst’ scar area, respectively. The rater pair R1 vs. R2 showed the greatest limits of agreement for total mVSS scores for both the ‘best’ and ‘worst’ scar area of −2.15 to 3.28 and −2.81 to 3.54, respectively. In Table 4, for the scar %TBSA estimation, the mean differences varied from −0.51 to 0.20. Rater pair R2 vs. R3 showed the greatest limits of agreement for scar %TBSA estimation of −3.99 to 2.97. With respect to allocation of the scar %TBSA category to <5 mVSS and 5–10 mVSS, the mean differences varied from −0.32 to 0.09 and −0.33 to 0.29, respectively. Rater pair R2 vs. R3 showed the greatest limits of agreement for allocation of scar %TBSA to <5 mVSS category (−3.82 to 3.19). Rater pair R1 vs. R2 had the greatest limits of agreement for allocation of scar %TBSA to 5–10 mVSS (−1.96 to 2.54).

Table 5 shows the clinically relevant misclassification rates of individual rater pairs. The vascularity component of the mVSS had differences of 2 or more units in less than 5% of score pairs in the ‘best’ and ‘worst’ scar area. The pliability component had differences 2 or more units in 2% of score pairs in the ‘best’ scar area and 8.9% of score pairs in the ‘worst’ scar area. There were no differences of 2 or more units in the height component for the ‘best’ or ‘worst’ scar area. For the total mVSS score there was misclassification (mVSS <5 versus ≥5) for 4.4% of score pairs in the ‘best’ scar area and 13.3% of score pairs in the ‘worst’ area of the scar. In terms of the allocation of scar %TBSA to severity categories (<5, 5–10, >10 mVSS), misclassifications of >1% scar TBSA occurred in 18.9% of score pairs in the <5 mVSS category and 7.8% of score pairs in the 5–10 mVSS category.

4. Discussion

The method proposed in this study links the mVSS (Baryza modification) to the %TBSA of the assessed scar. This enables scars to be monitored in terms of change in scar %TBSA in three mVSS categories. Using the mVSS alone, a selected area of the scar is assessed, and the maximum mVSS score may not change over the monitoring period. This study assessed the inter-rater reliability of the elements of the new mVSS-TBSA method, namely, mVSS scoring and %TBSA allocation in three categories of scar severity (<5 mVSS, 5–10 mVSS, >10 mVSS).

For the ‘worst’ scar area, there was ‘good to excellent’ inter-rater reliability for three of the individual mVSS components (vascularity, pliability and height), with ‘marginal’ reliability demonstrated for pigmentation assessment. ‘Good to excellent’ inter-rater reliability was shown for all individual mVSS components (pigmentation, vascularity, pliability and height) in ‘best’ area of the scar.

The assessment of pigmentation is highly subjective, whereby raters need to make judgements on the ‘browness’ of the scar. In addition, pigmentation is confounded by the

<table>
<thead>
<tr>
<th>Table 2 – Weighted Kappa ($k_w$) values for individual rater pairs in ‘best’ and ‘worst’ scar areas.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best area of the scar (n = 30)</td>
</tr>
<tr>
<td>$k_w$</td>
</tr>
<tr>
<td>Pigmentation</td>
</tr>
<tr>
<td>Vascularity</td>
</tr>
<tr>
<td>Pliability</td>
</tr>
<tr>
<td>Height</td>
</tr>
<tr>
<td>Worst area of the scar (n = 30)</td>
</tr>
<tr>
<td>$k_w$</td>
</tr>
<tr>
<td>Pigmentation</td>
</tr>
<tr>
<td>Vascularity</td>
</tr>
<tr>
<td>Pliability</td>
</tr>
<tr>
<td>Height</td>
</tr>
</tbody>
</table>

* nd = indeterminate as one observer showed no variation.

<table>
<thead>
<tr>
<th>Table 3 – ICC and Bland Altman values (mean score difference, 95% limits of agreement) for total mVSS for ‘best’ and ‘worst’ area of scar for rater pairs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 vs. R2</td>
</tr>
<tr>
<td>Total mVSS score (best scar area)</td>
</tr>
<tr>
<td>ICC (95% CI)</td>
</tr>
<tr>
<td>Mean difference (95% CI)</td>
</tr>
<tr>
<td>Lower limit of agreement (95% CI)</td>
</tr>
<tr>
<td>Upper limit of agreement (95% CI)</td>
</tr>
<tr>
<td>Proportion within limits of agreement</td>
</tr>
<tr>
<td>Total mVSS score (worst scar area)</td>
</tr>
<tr>
<td>ICC (95% CI)</td>
</tr>
<tr>
<td>Mean difference (95% CI)</td>
</tr>
<tr>
<td>Lower limit of agreement (95% CI)</td>
</tr>
<tr>
<td>Upper limit of agreement (95% CI)</td>
</tr>
<tr>
<td>Proportion within limits of agreement</td>
</tr>
</tbody>
</table>
colour of the scar due to vascularity even when ‘blanched’, since the degree of blanching depends on the pressure on the plexiglass by each rater. Forbes-Duchart et al. [21] also found that pigmentation was not a reliable parameter.

The $k_w$ score for height assessment in the ‘best’ area of the scar for two of the rater combinations was undetermined, a paradox that has been described in the literature [20,22,23]. This paradox reflects the low prevalence of rating 1 and high prevalence of rating 0 (normal skin) with a high observed proportion of agreement [22,23]. The other rater pair (R1 vs. R3) showed ‘excellent’ rater agreement for height assessment in the ‘best’ scar area. In the ‘worst’ scar areas, all rater combinations demonstrated ‘good to excellent’ rater agreement.

The inter-rater reliability of the total mVSS scores varied according to the area of the scar being tested: ‘best’ versus ‘worst’ scar area. Overall, higher levels of inter-rater agreement were observed for total mVSS measurements of the ‘worst’ area of the scar (‘good to excellent’ agreement) than the ‘best’ area (‘fair to good’ agreement). Clinically relevant misclassification (defined as <5 versus ≥5 mVSS) was infrequent in the ‘best’ scar area (4.4% of score pairs) but more frequent in the ‘worst’ scar area (13.3% of score pairs). Draaijers et al. [24] also showed that with three raters the ICC for total mVSS score using their modification of the VSS was ‘excellent’ (ICC 0.86–0.88 at 95% CI) and our findings for the inter-rater reliability for assessment of the ‘worst’ area of the scar are very similar (ICC 0.85–0.88). It is reasonable to assume that the area of scar assessed in Draaijers et al. [24] was the ‘worst’ scar area. The study presented here provides important new data on how the inter-rater reliability of the total mVSS score depends on the severity of the scar area being assessed.

Estimation of scar %TBSA demonstrated ‘excellent’ inter-rater agreement between all rater pairs but the agreement for allocation of scar %TBSA to mVSS severity categories was variable. Allocation of scar %TBSA to the <5 mVSS category was more reliable (‘good to excellent’) than the 5–10 category (‘fair to good’). We drew the corollary that the allocation of scar %TBSA into two complimentary categories (<5 and ≥5 mVSS) has ‘excellent to good’ reliability; however, from a clinical perspective, misclassifications of >1% scar TBSA in the <5 mVSS category occurred in 18.9% of score pairs.

A limitation of this study was that there were insufficient data to complete a meaningful analysis on mVSS scores over 10. Only three of the study sample of 30 subjects had mVSS scores in the category mVSS >10 according to one or more of the raters. Further assessments that incorporate greater numbers of subjects with areas of poor scar outcomes are required to establish the reliability of scar % TBSA allocation to >10 mVSS. The sample size chosen for this study was based on recommendations by Walter et al. [25]. However, a larger

---

### Table 4 – ICC and Bland Altman values (mean score difference, 95% limits of agreement) for scar %TBSA and allocation of scar %TBSA to <5 mVSS and 5–10 mVSS categories for rater pairs.

<table>
<thead>
<tr>
<th>% TBSA</th>
<th>R1 vs. R2</th>
<th>R2 vs. R3</th>
<th>R1 vs. R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5 mVSS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICC (95% CI)</td>
<td>0.95 (0.90–0.97)</td>
<td>0.01 (0.82–0.95)</td>
<td>0.96 (0.93–0.98)</td>
</tr>
<tr>
<td>Mean Difference (95% CI)</td>
<td>0.20 (−0.26, 0.65)</td>
<td>−0.51 (−1.16, 0.14)</td>
<td>−0.31 (−0.73, 0.10)</td>
</tr>
<tr>
<td>Lower limit of agreement (95% CI)</td>
<td>−2.26 (−3.05, −1.46)</td>
<td>−3.99 (−5.11, −2.86)</td>
<td>−2.53 (−3.23, −1.80)</td>
</tr>
<tr>
<td>Upper limit of agreement (95% CI)</td>
<td>2.65 (1.86, 3.44)</td>
<td>2.97 (1.84, 4.09)</td>
<td>1.89 (1.18, 2.60)</td>
</tr>
<tr>
<td>Proportion within limits of agreement</td>
<td>29/30 (97%)</td>
<td>29/30 (97%)</td>
<td>28/30 (93%)</td>
</tr>
<tr>
<td>5–10 mVSS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICC (95% CI)</td>
<td>0.80 (0.62–0.89)</td>
<td>0.78 (0.59–0.89)</td>
<td>0.63 (0.35–0.80)</td>
</tr>
<tr>
<td>Mean Difference (95% CI)</td>
<td>0.09 (0.49–0.67)</td>
<td>−0.32 (−1.01–0.38)</td>
<td>−0.23 (−0.90–0.43)</td>
</tr>
<tr>
<td>Lower limit of agreement (95% CI)</td>
<td>−2.82 (−3.82, −1.82)</td>
<td>−3.82 (−5.01, −2.62)</td>
<td>−3.51 (−4.66, −2.36)</td>
</tr>
<tr>
<td>Upper limit of agreement (95% CI)</td>
<td>3.00 (2.01, 4.00)</td>
<td>3.19 (1.99, 4.39)</td>
<td>3.05 (1.89, 4.19)</td>
</tr>
<tr>
<td>Proportion within limits of agreement</td>
<td>25/27 (93%)</td>
<td>25/27 (93%)</td>
<td>25/27 (93%)</td>
</tr>
</tbody>
</table>

### Table 5 – Clinically relevant misclassification rates of individual rater pairs.

<table>
<thead>
<tr>
<th>Pair wise misclassification ratesa</th>
<th>‘Best’ area of scar</th>
<th>‘Worst’ area of scar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vascularityb</td>
<td>4/90</td>
<td>3/90</td>
</tr>
<tr>
<td>Flabilityc</td>
<td>0/90</td>
<td>0/90</td>
</tr>
<tr>
<td>Heightb</td>
<td>2/90</td>
<td>8/90</td>
</tr>
<tr>
<td>Total mVSS scorec</td>
<td>0/90</td>
<td>0/90</td>
</tr>
<tr>
<td>For the whole scar</td>
<td>4/90</td>
<td>12/90</td>
</tr>
</tbody>
</table>

* Total pair-wise rater tests = 90 (3 assessments per patient for 30 patients).
* Difference of 2 units in the sub-scale score.
* Misclassification on basis of <5 and ≥5 mVSS.
* Misclassification on basis of >1% scar TBSA.

---

The inter-rater reliability of the total mVSS scores varied according to the area of the scar being tested: ‘best’ versus ‘worst’ scar area. Overall, higher levels of inter-rater agreement were observed for total mVSS measurements of the ‘worst’ area of the scar (‘good to excellent’ agreement) than the ‘best’ area (‘fair to good’ agreement). Clinically relevant misclassification (defined as <5 versus ≥5 mVSS) was infrequent in the ‘best’ scar area (4.4% of score pairs) but more frequent in the ‘worst’ scar area (13.3% of score pairs). Draaijers et al. [24] also showed that with three raters the ICC for total mVSS score using their modification of the VSS was ‘excellent’ (ICC 0.86–0.88 at 95% CI) and our findings for the inter-rater reliability for assessment of the ‘worst’ area of the scar are very similar (ICC 0.85–0.88). It is reasonable to assume that the area of scar assessed in Draaijers et al. [24] was the ‘worst’ scar area. The study presented here provides important new data on how the inter-rater reliability of the total mVSS score depends on the severity of the scar area being assessed.

Estimation of scar %TBSA demonstrated ‘excellent’ inter-rater agreement between all rater pairs but the agreement for allocation of scar %TBSA to mVSS severity categories was variable. Allocation of scar %TBSA to the <5 mVSS category was more reliable (‘good to excellent’) than the 5–10 category (‘fair to good’). We drew the corollary that the allocation of scar %TBSA into two complimentary categories (<5 and ≥5 mVSS) has ‘excellent to good’ reliability; however, from a clinical perspective, misclassifications of >1% scar TBSA in the <5 mVSS category occurred in 18.9% of score pairs.

A limitation of this study was that there were insufficient data to complete a meaningful analysis on mVSS scores over 10. Only three of the study sample of 30 subjects had mVSS scores in the category mVSS >10 according to one or more of the raters. Further assessments that incorporate greater numbers of subjects with areas of poor scar outcomes are required to establish the reliability of scar % TBSA allocation to >10 mVSS. The sample size chosen for this study was based on recommendations by Walter et al. [25]. However, a larger
sample size would have enabled determination of reliability estimations with greater statistical confidence [18]. Whilst the patient sample used in this study reflects the patient population of adults at our burns unit, a larger sample size may have provided more patients with diverse scar outcomes. The ICC values derived in this study suggest that the inter-rater reliability of the mVSS-TBSA method used is good. However, the results of the Bland Altman analyses revealed distributions of differences between rater scores that would question the level of reliability derived by the ICC analyses. Interpretation of the mean and limits of agreement of the difference in rater scores derived by the Bland Altman method relies on the clinical context, and acceptable clinical agreement cannot be based purely on statistics [20].

Clinically relevant misclassification rates of individual rater pairs were low for the ‘best’ and ‘worst’ areas of the scar for the components of the mVSS and the total mVSS score in the ‘best’ scar area. However, there was disagreement in classification of the total mVSS score for the ‘worst’ scar area in 13.3% of score pairs. Misclassifications of >1% scar TBSA occurred in 18.9% of score pairs in the <5 mVSS category. This indicates that despite high levels of reliability proposed for this method using conventional interpretations of inter-rater reliability scores, for the mVSS-TBSA method there is potential to misclassify scar outcome in some cases. For example, there are challenges in allocating the scar %TBSA in different severity categories because of the diverse physical characteristics and appearance of scar areas that have the same mVSS score. Refresher training may be important for optimum inter-rater reliability using this method.

The authors acknowledge that many of the components of the mVSS are ordered categorical variables and the validity of combining such scores to a total mVSS score is questionable. It is also debatable whether the total mVSS score can be regarded as a continuous measurement, as has been done in previous published papers. However, peer-reviewed publications of inter-rater-reliability of the VSS and its modifications are based on total VSS score to determine scar outcome. To be consistent with previous use and publications of the VSS we have reported mVSS total scores.

The original VSS, developed by Sullivan in 1990 [26] was designed to rate scars approximately 4 cm \( \times \) 4 cm in size and one year post burn capturing four scar parameters: pigmentation, vascularity, pliability and height [6,10,26]. Despite these specifications it is broadly utilised to score scars of varying level of size and maturity. Furthermore, the VSS in its original version, and the many modifications since then, fail to capture the clinically relevant variation in scar outcome within each scar. The mVSS-TBSA method described in this study was specifically developed to take this variation in scar outcome into consideration, based on scales that are already familiar to burns clinicians and researchers. Adding a spatial dimension to scar assessment enhances the currently available information from scar assessments for clinicians beyond an indication of the severity of the ‘worst’ scar area. The mVSS-TBSA method could be used with a variety of reliable scar assessment tools that are designed to monitor specific areas of scarring to provide a better overall perspective of scarring. In clinical practice it is often the situation that more than one person undertakes burn scar assessments, and the personnel involved in assessments have varying levels of experience. Hence, it is important to establish the inter-rater reliability of methods used for burn scar assessment.

5. Conclusion

This innovation linking the scar %TBSA to the mVSS provides a new method for identifying changes in scar outcome for improved clinical surveillance and decision-making, as well as for research purposes. The findings of this study highlight the necessity to interpret inter-rater reliability results in the respective clinical context. Using conventional interpretations of reliability scores and examination of clinically relevant disagreement rates in rater scores, these results support the reliability of the new mVSS-TBSA method, whilst acknowledging that there is potential for misclassification of scar outcome in some cases. Efforts to improve the inter-rater reliability of pigmentation and vascularity components using more objective tests, and regular refresher training, will further improve the inter-rater reliability of total mVSS score and have a positive impact on the utility of the mVSS-TBSA.

Conflict of interest statement

All authors declare no conflict of interest.

Acknowledgements

We acknowledge the support of the Wound Management Innovation Cooperative Research Centre, who funded this study. I wish to acknowledge Sugeech Ariyaratna for his contribution with statistical analysis as well as proof reading of this paper. I also wish to thank and acknowledge Delia Hendrie, my external supervisor for her general support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.burns.2013.01.014.

References

[5] Bloemen MCT, van Gerven MS, van der Wal MBA, Verhaegen PDM, Middelkoop E. An objective device for
Appendix B: Reliability of scar assessments performed with an integrated skin testing device – The DermaLab Combo® (published work)
Reliability of scar assessments performed with an integrated skin testing device – The DermaLab Combo®

T.U. Gankande a,*, J.M. Duke a, P.L. Danielsen a, b, H.M. Dejong a, c, F.M. Wood a, c, H.J. Wallace a

a Burn Injury Research Unit, School of Surgery, The University of Western Australia, Australia
b Department of Dermatology, Bispebjerg University Hospital, Denmark
c Burn Outcomes Centre, Royal Perth Hospital, Australia

A R T I C L E   I N F O

Article history:
Accepted 31 January 2014

Keywords:
Burns
Scar assessment
Objective scar measurements
Inter-rater reliability
Test–retest reliability

A B S T R A C T

Background: The DermaLab Combo® is a device with potential to make objective measurements of key scar components – pigmentation, vascularity, pliability and thickness. This study assessed the inter-rater and test–retest reliability of these measurements.

Method: Three raters performed scar assessments on thirty patients with burn scars using the DermaLab Combo®. Measurements of pigmentation, vascularity, pliability and thickness were made and intra-class correlation coefficients (ICC) were derived for inter-rater and test–retest reliability.

Results: Inter-rater reliability was found to be “excellent” in the ‘best’ and ‘worst’ areas of the index scar and normal skin for pigmentation (ICC: 0.94–0.98) and thickness (ICC: 0.86–0.96). Test–retest reliability was also “excellent” for pigmentation (ICC: 0.87–0.89) and thickness (ICC: 0.92–0.97) in all areas. Vascularity showed “good” to “excellent” inter-rater reliability (ICC: 0.66–0.84) in all areas however test–retest reliability was “low” (ICC: 0.29–0.42). Test–retest reliability was “excellent” for pliability (ICC: 0.76–0.91). Technical limitations were encountered making measurements in some scars for thickness, and in particular, pliability.

Conclusion: The DermaLab Combo® measured pigmentation, thickness and pliability with “excellent” reliability. If future studies provide protocols to improve test–retest reliability of vascularity measurements and obtain pliability measurements more successfully, the DermaLab Combo® will be valuable device for scar assessment.

C 2014 Elsevier Ltd and ISBI. All rights reserved.

1. Introduction

Scarring is a consequence of burn with the potential for long-term physical and psychological impacts on patients [1]. Systematic scar assessment enables clinicians to monitor scar progression and researchers to evaluate interventions and factors influencing scar outcome.

Current scar assessment methods used in clinical practice are subjective and may have low inter-rater reliability [2]. The...

* Corresponding author at: Burn Injury Research Unit, School of Surgery, M318, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia. Tel.: +61 08 6488 8772; fax: +61 08 6488 8580.
E-mail addresses: Uliya.Gankande@uwa.edu.au, ugankande@meddent.uwa.edu.au (T.U. Gankande).
http://dx.doi.org/10.1016/j.burns.2014.01.025
0305-4179/© 2014 Elsevier Ltd and ISBI. All rights reserved.
inter-rater reliability of the modified Vancouver Scar Scale (mVSS), for example, varies depending on which scar component is being measured [3]. Objective scar measurements are likely to provide more reliable data for clinical and research purposes, enabling more robust comparisons within and between patients [4–6].

Scar assessment includes measurement of four key components – pigmentation, vascularity, pliability and height (thickness) of the scar. Many individual devices have been tested to obtain objective measurements of these components, but none have been broadly adopted within a clinical setting. The most likely reason is that most devices only measure one individual component [7–9]. For example, in Oliveira and others [9], objective measurements for the four scar components were obtained using four separate devices.

Scar measuring devices can be broadly categorised into three groups, assessing colour (pigmentation and vascularity), pliability (elasticity) and height (thickness) [8]. To date no device capable of measuring all four scar components has been tested for reliability [7,8]. The DermaLab Combo® (Cortex Technologies, Denmark) [10] is a commercially available skin testing device that has the potential capability to measure all four scar assessment components. It was primarily designed for skin testing in the cosmetic field and does not claim to be a medical device. However, it is a high-sensitivity device based on proven technologies (see Section 2).

This study was designed to test the reliability (inter-rater reliability and test–retest reliability) of the DermaLab Combo® device to measure the four components of the mVSS in burn scars: pigmentation, vascularity, pliability and height (thickness) [11]. The study was conducted at Royal Perth Hospital (RPH) burns outpatient clinic, Western Australia.

2. Methods

2.1. Subjects

Thirty adult patients who attended the RPH burns outpatient clinic for routine follow-up assessments and were able to attend a follow-up appointment within 10 days of the first assessment were recruited consecutively. Participation was voluntary. Subjects were included in the study if they met the following inclusion criteria: at least 18 years of age; able to provide informed written consent; had a healed burn scar with an area of at least 3 cm x 6 cm; the scar was a minimum of three months post burn; and there was a contralateral, normal skin area of at least 3 cm x 3 cm with the same degree of sun exposure.

Subjects were excluded from the study if they could not provide informed written consent. If the subject wore a pressure garment, the rater responsible for recruiting the subject made a subjective assessment on the level of sun exposure of the contralateral normal skin and asked supplementary questions to confirm a similar level of sun exposure on both scar and contralateral normal skin areas.

All subjects were provided with a patient information sheet and a verbal explanation of the study after which written informed consent was sought.

2.2. Study design

The study was a single-arm observational study using three independent raters. For the inter-rater reliability study, all three raters made measurements of the scar and contralateral normal skin (control) areas for each parameter on the same day. For the test–retest reliability study, the same three raters made the same measurements on a second occasion within 10 days of the initial baseline assessment. The study was approved by Royal Perth Hospital (RPH) and University of Western Australia (UWA) Human Research Ethics Committees.

2.3. Raters

The three raters had comprehensive training in the use of the DermaLab Combo® under the supervision of an experienced user of the DermaLab Combo® (HW). Each training session was 3 h in duration and the raters undertook three sessions of training over a three week period. The training session consisted of the raters familiarising themselves with the use of the probes and the study protocol to measure the four components of scar assessment on each other. Subsequent to this training, each rater used the DermaLab Combo® device on patients with burn scars to conduct at least 10 independent assessments of scars using the study protocol prior to the start of the study. In this paper the raters are referred to as raters R1, R2 and R3.

2.4. Materials

The DermaLab Combo® was used to assess melanin (pigmentation), erythema (vascularity), elasticity (pliability) and thickness (height) of post-burn scars. The DermaLab Combo® consists of a main unit with screen and multiple separate probes. The three probes used in this study had separate channel connectors and were each attached to the main unit with separate tubes (Fig. 1).

2.4.1. The colour probe

The colour measurement of the DermaLab Combo® is based on the principle of narrow-band reflectance spectrophotometry (550 nm ± 30 nm and 660 nm ± 60 nm for haemoglobin [erythema or vascularity] and melanin [pigmentation] respectively). The colour probe has an optical focussing on 7 mm diameter target area with a clear front for accurate positioning and is illuminated by two angled white light emitting diode (LED) lights. The probe displays four readings (three individual measurements with their average) separately for erythema (vascularity) and melanin (pigmentation) using Commission Internationale de l’Eclairage (CIE) – luminance (L), red-green axis (a) and blue-yellow axis (b) [CIELab] values.

2.4.2. The elasticity probe

The elasticity measurement of the DermaLab Combo® is based on the principle of vertical suction applied on the surface of scar. The probe has a measuring aperture of 10 mm diameter and adheres to the skin by a double adhesive sticker. Elasticity measurements are expressed in terms of Young’s modulus in megaPascals (MPa). For the purpose of this paper, elasticity measurements will be referred to as ‘pliability’.
2.4.3. The skin thickness probe
The DermaLab Combo® skin thickness measurement probe is based on the principle of high frequency ultrasound (20 MHz) and has a resolution of 60 \( \mu \text{m} \times 200 \mu\text{m} \) with a penetration capacity of 3.4 mm with a fully adjustable gain settling (±10 dB). The probe has a rotating transducer, scan length 17 mm, footprint 11 mm, and the capacity to display actual and stored measurements side-by-side.

In contrast to the VSS height measurement that measures scar height above the pre-injury ‘baseline’ surface of the skin, this probe measures the thickness of the whole scar including the scar above and below the baseline surface of the skin. For the purpose of this paper, scar height measurements will be referred to as ‘thickness’.

### 2.5. Procedure

#### 2.5.1. Inter-rater reliability
All pressure garments and bandages were removed at least 15 min before the start of the assessment to reduce any blanching effect that may influence measurements. All assessments were carried out in the same room where the subjects were exposed to similar environmental conditions at the time of assessment. All assessments were carried out with the subjects in a sitting position. The anatomical body part containing the scar was placed in a nondependent position; if the scar was on a leg – both legs were elevated to rest on another chair, if the scar was on an arm – both arms were rested parallel to each other on a table.

One index scar of at least 6 cm \( \times \) 3 cm in size was identified per eligible participant. The index scar was defined as the scar with the largest percentage of body surface area on a body site (head/neck; chest; back; left arm; right arm; left leg; right leg) with a contralateral anatomically matched normal skin area available (minimum surface area 3 cm \( \times \) 3 cm). Within the index scar, one nominated rater selected a 3 cm \( \times \) 3 cm scar area deemed to be the ‘best’ area of the scar (lowest mVSS score) and a 3 cm \( \times \) 3 cm area of the same scar deemed to be the ‘worst’ area of the scar (highest mVSS score) using the mVSS [11]. These two areas were marked by this rater with a semi-permanent skin marker and photographed. A 3 cm \( \times \) 3 cm contralateral anatomically matched normal skin area was also identified and marked.

All three raters assessed these three marked areas using the DermaLab Combo®. Each scar component was assessed at three sites within the identified 3 cm \( \times \) 3 cm scar areas (see Fig. 2). These individual measurements as well as the average of the three measurements were recorded and used in the analysis. The three independent raters performed the scar assessments on each subject on the same day. In the case of pliability measurement, suction applied by the DermaLab Combo® elasticity probe alters the biomechanical properties of the skin, hence repeat measurements on the same location are not recommended by the manufacturer at less than 1 h intervals. Due to this extended wash-out period the elasticity probe was used only by one rater (R1) and pliability measurements were excluded from the inter-rater reliability analysis.

#### 2.5.2. Test–retest reliability
All study subjects were asked to attend a repeat assessment within 10 days of the initial baseline assessment. Two or three...
of the original raters performed the repeat DermaLab Combo
scar assessment on each subject. The repeat assessment was
carried out under the same conditions as the initial assess-
ment. Subjects were asked about anything that may have
affected their scar between the time of the initial baseline
assessment and the repeat assessment (for example, use of
different skin moisturiser, a new therapeutic measure such as
a different pressure garment or excessive sun exposure). If the
three marked scar areas were not visible they were re-marked
with the semi-permanent skin marker with the help of the
baseline photographs.

As for the initial baseline assessment, each scar parameter
was assessed on three sites within each marked area (Fig. 2)
and the individual measurements as well as the average of the
three measurements were recorded. The raters performed the
repeat scar assessments on each participant on the same day.
Due to the extended wash-out period (1 h) with the elasticity
probe, only rater 1 conducted the repeat pliability assessment.

3. Data collection and analysis

Data were collected over a 6-month period, from July 2012 to
January 2013. Each rater was blinded to the results of the other
raters and independently recorded results in a separate Scar
Assessment Form (Appendix 1). Each rater recorded four
measurements on the data collection sheet for each parameter
in the three 3 cm × 3 cm areas: one measurement for each site
and the average of the three measurements. All forms were filed
separately until the data collection was complete.

Throughout the paper we refer to the 3 cm × 3 cm squares
marking the ‘best’ and ‘worst’ parts of the index scar and
contralateral normal skin as “areas” and the probe placement
locations for measurements within the areas as “sites”.

For the purpose of the reliability analyses the measure-
ments for each site were considered as independent units for
analysis. Analyses were also conducted using the average
measurement of the three sites in the three 3 cm × 3 cm areas
for inter-rater reliability. All data analyses were performed
using SPSS version XX (Chicago, Inc) and Stata V12 statistical
software (StataCorp LP, College Station, TX).

3.1. Inter-rater reliability

Inter-rater reliability is the extent of agreement among raters
scoring the same subjects under the same conditions [12]. The
inter-rater reliability of individual measurements and the
average measurement for each component were calculated for
each rater pair (R1 vs. R2, R1 vs. R3, and R2 vs. R3) for the ‘best’
and ‘worst’ area of each index scar and the contralateral
normal skin area on each subject.

Measurements obtained with the DermaLab Combo device
were continuous variables. The Bland Altman (BA)
method and plots of the difference in measurements vs.
average of scores, for each rater pair were generated with
mean difference and 95% limits of agreement. BA plots
and results were examined for any patterns of variation to change
with the magnitude of the measurement. In the absence of
such patterns inter-rater reliability was established using two-
way random effects ANOVA to derive intra-class correlation
coefficients (ICC) with 95% confidence intervals (CI) for all
rater combinations. These methods are fully described in
Gankande and others [3]. ICC were interpreted using the
Rosner interpretation (0–0.40: marginal agreement; >0.40–
0.75: good agreement; >0.75: excellent agreement) [13].

3.2. Test–retest reliability

Test–retest reliability refers to the reliability or the consistency
of a measurement over time [14]. Repeat scar assessments
were done within 10 days of the initial baseline assessment,
during which time no significant change in scar was
anticipated. ICC were used as the measure of test-retest reliability [15,16]. Test-retest reliability of individual site measurements for each component was calculated for the ‘best’ and ‘worst’ area of each index scar and the contralateral normal skin area on each subject. As for inter-rater reliability, the BA method was applied and test-retest reliability was assessed using two-way random effects ANOVA to derive ICC with 95% confidence intervals. ICC were interpreted using the Rosner interpretation [13].

4. Results

4.1. Subjects and descriptive statistics
Of the 30 subjects, 12 (40%) were female and 18 (60%) were male. The median age of study subjects was 43 years (interquartile range [IQR]: 25.0–54.3 years; minimum (min)–maximum (max): 18–81 years). The median time since injury among subjects was 4.5 months (IQR: 3–6 months; min–max: 3–648 months). The locations of the assessed scars were: arm (n = 12), leg (n = 13), chest (n = 3), abdomen (n = 1) and back (n = 1). Twenty-seven out of the 30 subjects had a mVSS assessment performed. The median mVSS score in the ‘best’ area of scar was 2 (IQR: 1–2; min–max: 0–7) and 3 (IQR: 2–5; min–max: 1–10) in the ‘worst’ area of scar. The distributions of the scar component measurements at baseline for pigmentation, vascularity, thickness and pliability are described in Table 1.

4.2. Inter-rater reliability
ICC and 95% CI for all rater pairs (R1 vs. R2, R1 vs. R3 and R2 vs. R3) for pigmentation, vascularity and thickness measurements are presented for individual measurements (Table 2) and average measurements (Table 3).

Pigmentation had “excellent” inter-rater reliability in all the tested areas for all rater pairs, with ICC of 0.98, 0.94, 0.95 in the ‘best’ area of the scar, 0.96, 0.95, 0.97 in the ‘worst’ area of the scar and 0.95, 0.92, 0.91 in the contralateral normal skin area. When the average measurements for pigmentation were used in the analysis the inter-rater reliability was marginally greater among all rater combinations.

Vascularity showed “good” to “excellent” inter-rater reliability in all three areas measured. ICC of 0.74, 0.66 and 0.78 in the ‘best’ scar area and 0.84, 0.67 and 0.73 in the ‘worst’ scar area correspond to “good” to “excellent” inter-rater reliability. In the contralateral normal skin area the inter-rater reliability was only “good” and the ICC varied between rater pairs. Rater pairs R1 vs. R2 and R1 vs. R3 had higher inter-rater reliability (ICC 0.73) than the rater pair R2 vs. R3 (ICC 0.54). As

<table>
<thead>
<tr>
<th>Component</th>
<th>Minimum measurement</th>
<th>Maximum measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pigmentation</td>
<td>26.00</td>
<td>93.00</td>
</tr>
<tr>
<td>Vascularity</td>
<td>10.00</td>
<td>45.00</td>
</tr>
<tr>
<td>Thickness</td>
<td>1001.00</td>
<td>2109.00</td>
</tr>
<tr>
<td>Pliability</td>
<td>177.00</td>
<td>3240.00</td>
</tr>
<tr>
<td>Normal skin area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation</td>
<td>28.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Vascularity</td>
<td>7.00</td>
<td>51.00</td>
</tr>
<tr>
<td>Thickness</td>
<td>1132.0</td>
<td>2433.00</td>
</tr>
<tr>
<td>Pliability</td>
<td>264.00</td>
<td>9438.00</td>
</tr>
</tbody>
</table>

Table 2 – Inter-rater reliability for pigmentation, vascularity and thickness components based on individual site measurements (all rater combinations).

<table>
<thead>
<tr>
<th>Component</th>
<th>R1 vs. R2&lt;sup&gt;a&lt;/sup&gt;</th>
<th>R1 vs. R3&lt;sup&gt;b&lt;/sup&gt;</th>
<th>R2 vs. R3&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Best' scar area</td>
<td>ICC (95% CI)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>ICC (95% CI)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>ICC (95% CI)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Pigmentation</td>
<td>0.98 (0.96, 0.98)</td>
<td>0.94 (0.90, 0.96)</td>
<td>0.95 (0.92, 0.98)</td>
</tr>
<tr>
<td>Vascularity</td>
<td>0.74 (0.60, 0.83)</td>
<td>0.66 (0.48, 0.79)</td>
<td>0.78 (0.66, 0.85)</td>
</tr>
<tr>
<td>Thickness</td>
<td>0.95 (0.92, 0.96)</td>
<td>0.86 (0.78, 0.91)</td>
<td>0.93 (0.89, 0.95)</td>
</tr>
<tr>
<td>'Worst' scar area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation</td>
<td>0.96 (0.94, 0.98)</td>
<td>0.95 (0.93, 0.97)</td>
<td>0.97 (0.95, 0.98)</td>
</tr>
<tr>
<td>Vascularity</td>
<td>0.84 (0.76, 0.89)</td>
<td>0.67 (0.50, 0.78)</td>
<td>0.73 (0.59, 0.82)</td>
</tr>
<tr>
<td>Thickness</td>
<td>0.95 (0.91, 0.97)</td>
<td>0.91 (0.85, 0.95)</td>
<td>0.96 (0.92, 0.97)</td>
</tr>
<tr>
<td>Normal skin area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation</td>
<td>0.95 (0.93, 0.97)</td>
<td>0.92 (0.87, 0.95)</td>
<td>0.91 (0.86, 0.94)</td>
</tr>
<tr>
<td>Vascularity</td>
<td>0.73 (0.58, 0.82)</td>
<td>0.73 (0.58, 0.82)</td>
<td>0.54 (0.30, 0.69)</td>
</tr>
<tr>
<td>Thickness</td>
<td>0.94 (0.91, 0.96)</td>
<td>0.92 (0.88, 0.95)</td>
<td>0.95 (0.92, 0.97)</td>
</tr>
</tbody>
</table>

<sup>a</sup> ICC (95% CI): intra-class correlation coefficient (95% confidence interval).
<sup>b</sup> Raters 1–3.
<sup>c</sup> 3 site measurements for one subject were not included in the analysis (reached maximum measurement).
<sup>d</sup> 33 site measurements not included in the analysis (reached maximum measurement).
with pigmentation, there was marginally higher inter-rater reliability for vascularity when the average measurement was used in the analysis.

For about one-third of the 30 subjects (11/30), the thickness measurement in the ‘worst’ area of scar reached the maximum thickness of 2.5 mm. Only one of the 30 subjects reached the maximum thickness in the ‘best’ area of the scar. When thickness was able to be measured, it achieved “excellent” inter-rater reliability in all the tested areas for all rater pairs with ICC values of 0.95, 0.86 and 0.93 in the ‘best’ area of scar, 0.95, 0.91 and 0.96 in the ‘worst’ area of scar and 0.94, 0.92 and 0.95 in the contralateral normal skin area. Again, there was marginally higher inter-rater reliability when the average measurement was used in the analysis.

4.3. Test–retest reliability

Repeat scar assessments were completed for 19 out of the 30 subjects. For measurement of pigmentation, vascularity and thickness, 12 subjects had repeat testing conducted by three raters while 7 were assessed by only two raters, resulting in a total of 50 repeat tests for analysis. As per the protocol, each rater took measurements at 3 sites of each area of assessment (‘best’ and ‘worst’ area of index scar and the contralateral normal skin area). For analysis of individual measurements (3 sites per area in 50 repeat tests), up to 150 paired data sets were available for test–retest analysis. All subjects confirmed to the rater team that nothing had affected their scar since the first assessment including different topical preparations, pressure therapy or excessive sun exposure.

The results of the test–retest reliability ICC and 95% CI for pigmentation, vascularity and thickness are presented in Table 4. The number of data pairs included in each analysis is indicated.

Pigmentation had “excellent” test–retest reliability in the ‘best’ and ‘worst’ areas of the scar and the contralateral normal skin area (ICC 0.87, 0.89, 0.83 respectively).

Vascularity demonstrated varying levels of test–retest reliability. In the ‘worst’ area of the scar test–retest reliability for vascularity was “good” (ICC 0.42). In the ‘best’ area of scar and the contralateral normal skin area test–retest reliability was “marginal” with ICC 0.29 and 0.39 respectively.

Thickness also showed “excellent” test–retest reliability in all the three areas: the ‘best’ area of the scar (ICC 0.97), ‘worst’ area of the scar (ICC 0.92) and the contralateral normal skin area (ICC 0.86).

### Table 3 – Inter-rater reliability for pigmentation, vascularity and thickness components using average measurements of each area (all rater combinations).

<table>
<thead>
<tr>
<th>Component</th>
<th>R1 vs. R2&lt;sup&gt;a&lt;/sup&gt; ICC (95% CI)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>R1 vs. R3&lt;sup&gt;a&lt;/sup&gt; ICC (95% CI)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>R2 vs. R3&lt;sup&gt;a&lt;/sup&gt; ICC (95% CI)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Best’ scar area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation (n = 30)</td>
<td>0.98 (0.96, 0.99)</td>
<td>0.94 (0.88, 0.97)</td>
<td>0.95 (0.92, 0.98)</td>
</tr>
<tr>
<td>Vascularity (n = 30)</td>
<td>0.77 (0.52, 0.89)</td>
<td>0.68 (0.33, 0.85)</td>
<td>0.79 (0.56, 0.90)</td>
</tr>
<tr>
<td>Thickness (n = 29)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.97 (0.94, 0.99)</td>
<td>0.90 (0.77, 0.95)</td>
<td>0.95 (0.88, 0.96)</td>
</tr>
<tr>
<td>‘Worst’ scar area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation (n = 30)</td>
<td>0.97 (0.94, 0.99)</td>
<td>0.97 (0.93, 0.98)</td>
<td>0.97 (0.95, 0.99)</td>
</tr>
<tr>
<td>Vascularity (n = 30)</td>
<td>0.85 (0.69, 0.93)</td>
<td>0.68 (0.32, 0.85)</td>
<td>0.73 (0.43, 0.87)</td>
</tr>
<tr>
<td>Thickness (n = 19)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.98 (0.96, 0.99)</td>
<td>0.97 (0.91, 0.98)</td>
<td>0.98 (0.94, 0.99)</td>
</tr>
<tr>
<td>Normal skin area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigmentation (n = 30)</td>
<td>0.96 (0.92, 0.98)</td>
<td>0.93 (0.84, 0.96)</td>
<td>0.92 (0.82, 0.96)</td>
</tr>
<tr>
<td>Vascularity (n = 30)</td>
<td>0.72 (0.42, 0.87)</td>
<td>0.74 (0.44, 0.87)</td>
<td>0.57 (0.10, 0.79)</td>
</tr>
<tr>
<td>Thickness (n = 30)</td>
<td>0.97 (0.93, 0.98)</td>
<td>0.96 (0.92, 0.98)</td>
<td>0.98 (0.97, 0.99)</td>
</tr>
</tbody>
</table>

<sup>a</sup> ICC (95% CI): intra-class correlation coefficient (95% confidence interval).

<sup>b</sup> 1 subject not included in the analysis (reached maximum measurement).

<sup>c</sup> 11 subjects not included in the analysis (reached maximum measurement).

### Table 4 – Test–retest reliability for pigmentation, vascularity and thickness components using individual site measurements.

<table>
<thead>
<tr>
<th>Measurement area</th>
<th>Pigmentation (95% CI)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Vascularity (95% CI)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Thickness (95% CI)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Best’ area of scar</td>
<td>0.87 (0.82, 0.90) (n = 150)</td>
<td>0.29 (0.01, 0.48) (n = 150)</td>
<td>0.97 (0.89, 0.94) (n = 150)</td>
</tr>
<tr>
<td>‘Worst’ area of scar</td>
<td>0.89 (0.83, 0.92) (n = 150)</td>
<td>0.42 (0.19, 0.58) (n = 150)</td>
<td>0.92 (0.88, 0.95) (n = 99)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Normal skin area</td>
<td>0.83 (0.78, 0.88) (n = 150)</td>
<td>0.39 (0.15, 0.56) (n = 150)</td>
<td>0.86 (0.81, 0.89) (n = 150)</td>
</tr>
</tbody>
</table>

<sup>a</sup> ICC (95% CI): intra-class correlation coefficient (95% confidence interval).

<sup>b</sup> n = number of test–retest data pairs included in analysis.

<sup>c</sup> 51 test–retest data pairs not included in the analysis (at least one measurement reached maximum limit).
For pliability, difficulties in obtaining measurements were experienced and limited the number of test–retest data pairs for analysis. Matched measurements for test–retest analysis were not successfully obtained for 6 of the 19 subjects at any of the 3 sites in the ‘best’ area of the scar and the normal skin area. In the ‘worst’ area of the scar, matched measurements for test–retest analysis were not successfully obtained for 10 of the 19 subjects at any of the 3 sites. Particular difficulty obtaining successful measurements was experienced in the ‘worst’ area of the scar, on the legs and on body areas with thick hair. Results of test–retest reliability for pliability using available matched data were “excellent” in the ‘best’ (ICC 0.91, 95%CI: 0.78–0.96) and ‘worst’ (ICC 0.76, 95%CI: 0.29–0.92) areas of scar. However, in the contralateral normal skin area pliability only achieved “good” test–retest reliability (ICC 0.45, 95%CI: 0.30–0.76).

5. Discussion

The main purpose of this study was to measure scar components analogous to those measured in the subjective Vancouver Scar Scale (VSS), namely, pigmentation (melanin), vascularity (erythema), pliability (elasticity) and height (thickness), using the DermaLab Combo® device. These objective measurements were analysed to assess the inter-rater and test–retest reliability and the capability of the DermaLab Combo® device in conducting burn scar assessments.

The DermaLab Combo® is a relatively new commercially available integrated skin testing device primarily designed for skin testing in the cosmetic field. The advantages of this device are that it is user-friendly and measures multiple components, including those assessed in the VSS. Therefore, there is potential for this device to be used for clinical and research purposes if it is shown to be reliable in measuring scar components.

The narrow measuring head aperture of the DermaLab Combo® may be a source of potential bias when measuring large scars. A series of measurements are needed to obtain an average score representative of the entire scar [17]. The current study recorded and used three individual site measurements within 3 cm × 3 cm areas representing the ‘best’ and ‘worst’ areas of the index scar and contralateral normal skin. The individual site measurements and average of the three site measurements were both used for the purpose of analysis.

The results of this study show “excellent” inter-rater and test–retest reliability for DermaLab Combo® measurements of pigmentation. This is a marked improvement compared to the modified VSS (mVSS) assessment of pigmentation in our previous study [3] that demonstrated only “marginal” to “good” inter-rater reliability (weighted Kappa [kw] 0.02–0.33) in the ‘worst’ area of scar and “good” inter-rater reliability (kw 0.45–0.57) in the ‘best’ area of scar. No technical issues were encountered making pigmentation measurements.

Vascularity assessment indicated “good” to “excellent” inter-rater reliability in all scar areas. This represents a marginal improvement compared to the inter-rater reliability of the mVSS reported for the ‘best’ area of the scar in our previous study; “good” to “excellent” inter-rater reliability (kw 0.64–0.76) in the ‘worst’ area of the scar, and “good” inter-rater reliability (kw 0.44–0.71) in the ‘best’ area of the scar [3]. However, test–retest reliability for the vascularity component using the DermaLab Combo® was low and failed to achieve an acceptable level of reliability.

The inter-rater reliability of the vascularity measurements was not as high as we had expected, and test–retest reliability was particularly poor. This was in sharp contrast to the “excellent” reliability of the pigmentation measurements. Spectrophotometry, the measuring principle used in the DermaLab Combo® has been in use for over 50 years and is considered to be a reliable and objective method for skin colour assessment [18,19]. Compared to the human observer, reflectance spectrophotometry can detect very small changes in vascularity (erythema) or pigmentation [20]. This sensitivity may have been responsible for the lower than expected reliability for vascularity. The protocol of the current study did not specify with what pressure the probe should be applied and excessive pressure may cause blanching of the skin. Conversely, the measurement of pliability with the suction probe within a short space of time before erythema measurement may have increased the degree of erythema. These sources of variation require investigation in future studies, and could be avoided by modifications to the measurement protocol. Therefore, we recommend that protocols of future studies of scar assessment use the colour probe first followed by the ultrasound probe and finally the pliability probe to avoid possible influence on the other measurements. In addition, changes in the skin chromophore concentrations (melanin and haemoglobin) induce changes in both the melanin index (MI) and the erythema index (EI) in narrow-band spectrophotometry, making it difficult to separate the contribution of each component [18]. This may cause problems for the measurement of skin pigmentation and vascularity after an intervention (e.g. after suction), or over time (e.g. test–retest) where one or both skin chromophores on a location may be altered [18]. Calibration of the colour probe was repeated each day of testing as recommended by the manufacturer, but any issues in the calibration process would also manifest as poor test–retest reliability.

Thickness measurements achieved “excellent” inter-rater and test–retest reliability across all measured areas. Van der Kerckhove and others [21] used a DermaScan C® (high frequency ultrasound, Cortex Technologies, Denmark) [10] to measure scar thickness, and also obtained “excellent” results for inter-rater reliability (ICC 0.88). The DermaLab Combo® device shows superior inter-rater reliability for measuring scar thickness compared to the subjective mVSS which demonstrated only “good” to “excellent” inter-rater reliability in both the ‘worst’ and ‘best’ areas of the scar (kw 0.72–0.76) [3].

Despite achieving “excellent” reliability (both inter-rater and test–retest) of the thickness measurement, a technical limitation was encountered regarding the maximum thickness measurement using the DermaLab Combo®. The manufacturer’s specifications for the DermaLab Combo® ultrasound probe indicate a 3.4 mm penetration capacity. However, the maximum scar thickness that could be measured during our study was 2.5 mm. Approximately one-third of the 30 study subjects had readings in the ‘worst’ area of the scar that reached the maximum measurement. The authors understand an additional ultrasound probe is currently in development which may address this limitation.

The test–retest reliability of pliability using the DermaLab Combo® was “excellent” in the scar areas, but only “good” in
normal skin. In this study we did not assess inter-rater reliability of pliability, but Anthonissen and others [22] using a similar device, the DermaLab™ (Cortex Technologies, Denmark), found that inter-rater reliability of pliability of burn scars to be “excellent” for both scars and normal skin (grafted scar: ICC 0.86; spontaneously healed skin and normal skin: ICC 0.93), similar to the inter-rater reliability of pliability measurements made with the mVSS (‘best’ area of scar: kw 0.82–0.84; ‘worst’ area of scar: kw 0.77–0.86) [3]. A direct comparison between the DermaLab Combo® and the DermaLab™ studies is not possible because the two studies measured different types of reliability and the devices were not identical.

Despite the “excellent” test-retest reliability of pliability measurements with the elasticity (pliability) probe in scar areas, the probe had limitations. The raters found scar assessment with the probe was not achievable on approximately half of the ‘worst’ scar areas, suggesting the probe may not have appropriate specifications for making measurements on rigid tissue. Anthonissen and others [22] also discuss the potential limitations of the elasticity (pliability) probe of the DermaLab™ device to measure rigid scars. Difficulties were also observed obtaining successful measurements on the legs and we speculate that this may due to the greater thickness of the epidermis of the leg skin. In addition, the probe was ineffective in obtaining a proper grip with the skin with its adhesive tape when the subjects had a rough growth of body hair. The influence of body hair may help explain the low test-retest reliability for normal skin, as normal skin has more hair than scar. Protocols of future studies should include shaving the area prior to use of the elasticity probe.

A major strength of this study is the identification and systematic evaluation of a device with the potential capability of measuring four scar components considered important in scar assessment in an objective manner. This study has translated the DermaLab Combo® device, currently used in the cosmetic industry, to the area of burn scar assessment with some success. Pigmentation measurement was highly reliable, showing a substantial improvement compared to the mVSS assessment of pigmentation, and no technical issues were encountered. Scar thickness measurement was also highly reliable; however, for about one-third of the scars measured the thickness reached the maximum capacity of the device in its current version. Pliability measurement offered only a modest improvement in inter-rater reliability compared to the mVSS and was limited in its capacity to make measurements in approximately half of the burns scars tested. We are of the view that the vascularity assessment using the DermaLab Combo® did not reach its full potential in the current study.

6. Conclusion

The DermaLab Combo® is an easy to use and commercially available device, making it a viable option for scar assessment in both clinical and research settings. It provides objective and reliable measurements for pigmentation, thickness and pliability. However, the device has limitations in making measurements of thickness and pliability in some burns scars. Thickness can be measured in all scars but reaches a maximum measurement. The problems encountered obtaining successful measurements of pliability and reliable vascularity measurements are being examined further. If future studies provide protocols to improve test-retest reliability of vascularity measurements and obtain pliability measurements more successfully, the DermaLab Combo® will be a valuable option for scar assessment and monitoring.

Conflict of interest statement

All authors declare no conflict of interest.

Acknowledgements

The authors acknowledge the funding support of the Wound Management Innovation Cooperative Research Centre and Sugeech Ariyaratna for his contribution in proof reading of this paper.

Appendix 1

ROYAL PERTH HOSPITAL BURNS UNIT

DERMALAB COMBO - SCAR ASSESSMENT

<table>
<thead>
<tr>
<th>Colour</th>
<th>Elasticity (PD only)</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanin</td>
<td>Best</td>
<td>Best</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>Worst</td>
<td>Worst</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>Control</td>
<td>Control</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

REFERENCES


Appendix C: Interpretation of the DermaLab Combo® pigmentation and vascularity measurements in burn scar assessment: An exploratory analysis (accepted for publication on 20 January 2015)
Interpretation of the DermaLab Combo® pigmentation and vascularity measurements in burn scar assessment: An exploratory analysis


Burn Injury Research Unit, Burn Injury Research Unit, School of Surgery, M318, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

Burn Injury Research Unit, School of Surgery, The University of Western Australia, Australia

Burn Outcomes Centre, Royal Perth Hospital, Australia

Article Info

Article history:
Accepted 18 January 2015

Keywords:
Burn scars
Burn scar assessment
Objective burn scar measurements
Scar metrics
Pigmentation
Vascularity

Abstract

Background: The DermaLab Combo® measures pigmentation and vascularity of a burn scar more reliably than the modified Vancouver Scar Scale (mVSS). This study aims to examine how the DermaLab Combo® continuous measurements of pigmentation and vascularity of burns scars relate to the mVSS, a standard clinical scar assessment method; and secondly, to obtain evidence to support the concurrent validity of DermaLab Combo® measurements for pigmentation and vascularity.

Method: Scar assessments were performed on an index burn scar of 100 patients using two methods: the mVSS (two raters) and the DermaLab Combo® device (one rater). Using the DermaLab Combo® measurements of pigmentation and vascularity for the index scar and an adjacent normal skin site were obtained. Indices were generated to represent the scar pigmentation (melanin index, MI%) and scar vascularity (erythema index, EI%) relative to the patient’s matched normal skin. Exploratory univariate and bivariate analyses were conducted and the concordance of classification by mVSS score using DermaLab® cut-off values was assessed.

Results: For pigmentation, the results suggest a 80% classification concordance for the DermaLab Combo® MI% values into mVSS pigmentation categories (hypopigmentation, normal pigmentation and hyperpigmentation) using two predictors (MI% and EI%) and visually fitted discriminant axis cut-offs. Due to the high degree of overlap of EI% values between the vascularity categories, meaningful classification of EI% values using the mVSS was not possible.

Conclusion: Quantifying percentage changes in melanin and erythema relative to matched normal skin improved understanding of the DermaLab Combo® pigmentation and vascularity measurements. The DermaLab Combo® pigmentation MI% values were able to be classified into pigmentation categories of the mVSS, and pigmentation classification concordance was further improved with consideration of the scar’s DermaLab Combo® vascularity EI% values. The DermaLab Combo® is an objective tool; however, while the measurement provides continuous numerical data that may be useful for identifying change over time in clinical scar monitoring of pigmentation and vascularity, further work is required to understand the DermaLab Combo® measurements to optimise the interpretation of these data.

© 2015 Published by Elsevier Ltd and ISBI.
1. Introduction

Scar assessment is essential for diagnosis, monitoring, evaluation and therapeutic management of scar pathology [1]. Colour evaluation has remained problematic even with advances in the area of scar assessment scales [2,3]. Inter-rater reliability of subjective scales in evaluating colour has been shown to be poor [2,4] and requires multiple observers to achieve reliable assessment [2,5]. In the clinical setting, scars are typically assessed by a single observer and frequently by many different observers over time, further weakening the accuracy of subjective assessment.

The Vancouver Scar Scale (VSS) is a commonly used scar assessment method in many burns care facilities because it is inexpensive and relatively easy to perform. The VSS is a subjective method and measures four scar parameters: pigmentation, vascularity, pliability and thickness. A modified VSS [6] is currently used at the Royal Perth Hospital (RPH) in Western Australia to conduct scar assessments. In a previous study the inter-rater reliability of this method was established [7]. For pigmentation inter-rater reliability was ‘marginal’ in both ‘best’ (weighted kappa $\kappa_w = 0.5–0.6$) and ‘worst’ ($\kappa_w = 0.1–0.3$) areas of the scar while for vascularity inter-rater reliability was ‘good to excellent’ (in both ‘best’ $\kappa_w = 0.4–0.7$) and ‘worst’ ($\kappa_w = 0.6–0.8$) areas of the scar. This study made recommendations for improving the inter-rater reliability of pigmentation and vascularity components by using more objective tests [7].

The DermaLab Combo™ (Cortex Technologies, Denmark) is a commercially-available skin testing device with the capability to measure scar colour (pigmentation and vascularity), pliability and thickness. It was primarily designed for skin testing in the cosmetic field and does not claim to be a medical device. However, it is a high-specification device based on proven technologies. The DermaLab Combo™ measures skin colour by narrow-band spectrophotometry and provides continuous scar measurements for pigmentation (melanin index) and vascularity (erythema index). Our previous work compared the reliability of subjective scar assessments of the mVSS and objective measurements made with the DermaLab Combo™. The results showed that the DermaLab Combo™ measures pigmentation and vascularity of a scar more reliably than the mVSS [8].

For a diagnostic tool to have clinical utility it must demonstrate acceptable validity and reliability. Validity refers to evidence that supports that an instrument or tool is actually measuring what it says it is measuring [9]. Concurrent validity in the context of diagnostic instruments, is tested by comparing results obtained with a “test” instrument with those of a gold standard [9,10]. However, a gold standard assessment tool is currently not available for assessing pigmentation and vascularity of scars [11]. In the absence of a gold standard for validation of the DermaLab Combo™ device this study used the mVSS [6] as a reference standard. The study was conducted at the Royal Perth Hospital (RPH) burns outpatient clinic, Western Australia. This study aims firstly, to examine how the DermaLab Combo™ continuous measurements of pigmentation and vascularity of burns scars relate to the mVSS, the current standard clinical scar assessment method used in many burns units; and secondly, to obtain evidence to support the concurrent validity of DermaLab Combo™ measurements for pigmentation and vascularity.

2. Method

2.1. Subjects

Adult patients scheduled for regular scar assessments at the RPH burns outpatient clinic as part of routine follow-up care for a burn injury were targeted for recruitment. Participation was voluntary and all patients received the same standard of clinical care whether they elected to participate or not. The subjects were included in the study if they met the following inclusion criteria: at least 18 years of age; able to provide voluntary written informed consent; had a fully epithelialized burn scar with an area of at least $3 \text{ cm} \times 3 \text{ cm}$; and an adjacent normal skin area of at least $3 \text{ cm} \times 3 \text{ cm}$ with similar sun exposure. Subjects were excluded from the study if they were unable to provide informed written consent. All eligible subjects were provided with a patient information sheet and a verbal explanation of the study and written consent was sought.

2.2. Study design

The study was a single arm observational study with three scar raters. The three raters were independent and blinded to the results of each other. The study was approved by the RPH and University of Western Australia (UWA) Human Research Ethics Committees.

2.3. Raters

The three raters used in the study were all experienced scar assessors and included a senior occupational therapist (OT) with over 7 years’ experience in the mVSS method (stratifying rater-R1); a clinical researcher with over two years’ experience in scar assessment and in the use of the DermaLab Combo™ device (DermaLab Combo™ rater – R2); and a clinician expert rater with over 10 years’ experience in the mVSS method (burns surgeon-R3). For the remainder of this paper, raters will be referred to as R1, R2 and R3.

2.4. Scar assessment

2.4.1. The mVSS (Baryza and Baryza modification)

The mVSS [6] was used to assess pigmentation and vascularity of the index scar area and the adjacent normal skin area. This mVSS methodology applied has been fully described in our previous publication [7].

2.4.2. The DermaLab Combo™ device – colour probe

The colour measurement of the DermaLab Combo™ device (Fig. 1) is based on the principle of narrow-band reflectance spectrophotometry (550 nm $\pm$ 30 nm and 660 nm $\pm$ 60 nm for haemoglobin [erythema or vascularity] and melanin [pigmentation] respectively). The colour probe specifications has been fully described in our previous publication [8]. The unit displays four readings (three individual measurements and the average of the three measurements) separately for
2.5. Data collection and procedure

Data were collected over a 4-month period, from November 2013 to February 2014. Each rater was blinded to the results of the other raters and independently recorded results in scar assessment forms. All scar assessment forms were filed separately until the data collection was complete.

Study subjects were exposed to constant environmental conditions at the time of the assessments and all assessments were performed with the subjects in a sitting position. The anatomical body part containing the scar was placed in a nondependent position; if the scar was on a leg, both legs were elevated to rest on another chair, if the scar was on an arm, both arms were rested parallel to each other on a table. All pressure garments and bandages were removed at least 15 min prior to the initial assessment. A minimum 10 min wash-out period was observed between each rater. Each subject had three scar assessments performed within 1 h on the same day by all three raters.

The stratifying rater (R1) identified and marked with a semi-permanent marker, 3 cm × 3 cm areas on both the index scar (based on the mVSS) and an adjacent normal skin area [12]. Subjects were stratified by pigmentation categories (normal pigmentation, hypopigmentation, mixed pigmentation and hyperpigmentation) and vascularity categories (normal vascularity, pink, red and purple) during recruitment until each category achieved a minimum of 20 subjects. The data were recorded in the mVSS data collection sheet [7] and the data stratification sheet (Appendix I).

The DermaLab Combo® scar assessment (R2) was conducted on the marked index scar area and the adjacent normal skin area using the DermaLab Combo® colour probe. Within the index scar, R2 assessed three individual sites for pigmentation and vascularity (Fig. 2). The three measurements and the automatically generated average measurement were recorded on a scar assessment form (Appendix II). Fig. 2 shows the probe placement for the DermaLab Combo® measurement within both the two 3 cm × 3 cm measurement areas (index scar and the adjacent normal skin). The circles numbered 1, 2 and 3 represent the locations for each site measurement (probe placement). The locations themselves were not marked, as this would interfere with the readings.

Pigmentation and vascularity parameters were scored by clinical expert rater (R3) using the mVSS for the marked index scar and the scores were recorded in the scar assessment sheet (Appendix II).

3. Analysis and sample size

Descriptive statistics for age, gender, Fitzpatrick skin type and anatomical location of scar were performed. The DermaLab Combo®
Combọ®-generated average measurements were used to derive study variables for scar pigmentation and vascularity relative to the normal skin measurements: melanin index (M1%) (pigmentation) and erythema index (E1%) (vascularity).

3.1. Generation of melanin index (M1%) and erythema index (E1%)

After personal communication with the manufacturer of the DermaLab Combo® relative percentage changes in melanin and erythema were quantified. The percentage change in pigmentation ($\Delta P$) was calculated as the difference between the DermaLab Combo® pigmentation measurements for scar (Ps) and normal skin (Pn) divided by the normal skin measurements (Pn) and multiplied by 100 ($\Delta P = (Ps - Pn)/ Pn \times 100$ [%]). The following formula was used to generate the variable, melanin index (M1%) with normal skin pigmentation equated to 100%: $M1\% = 100 + \Delta P$ (%). Similarly, the percentage change in vascularity ($\Delta V$) was calculated as the difference between the DermaLab Combo® vascularity measurement for the scar (Vs) and the normal skin (Vn) divided by the normal skin measurement (Vn) and multiplied by 100 ($\Delta V = (Vs - Vn)/ Vn \times 100$ [%]). The following formula was then used to generate the variable, erythema index (E1%), with normal skin vascularity equated to 100%: $E1\% = 100 + \Delta V$ (%).

3.2. Exploratory data analysis

Data were graphed and visualised, descriptive statistics generated (median; inter quartile range [IQR]; minimum and maximum) and outliers identified for the variables M1% and E1% across the mVSS categories of pigmentation (normal pigmentation, hypopigmentation and hyperpigmentation) and vascularity (normal, pink, red, purple). Kruskal–Wallis test with post-test comparisons for non-parametric distributions were conducted. Dot plots with summary measures (median, minimum–maximum) were generated for pigmentation M1% and vascularity E1% by clinical classification using the mVSS.

To explore if the two variables were more useful than a single variable for discriminating mVSS categories, bivariate plots were generated and analyses of M1% and E1% data by mVSS categories were undertaken [13,14]. A linear discriminant axis that optimised separation of mVSS categories was identified using an iterative process from the bivariate plot. A discriminant value for each observation was calculated from both variables using the formula of the line perpendicular to the discriminant axis which passes through the observation: discriminant value (pigmentation) = $a_1 M1\% + b_1 E1\%$, discriminant value (vascularity) = $a_2 E1\% + b_2 M1\%$. Cut-off discriminant values for classification into mVSS categories were identified. An iterative process was used to identify discriminant cut-offs that optimised sensitivity and specificity. The concordance of classification into mVSS categories using two predictor variables was compared to results from using cut-off values for the single variable (M1% or E1%).

To assess the associations between the M1% and E1% derived values from the DermaLab Combo® measurements (continuous) with the respective coded numerical scores of mVSS (ordinal) pigmentation and vascularity, Kendall tau-b ($\tau_b$) rank correlations (accounting for ties) were performed [15].

In the absence of a gold standard, the subjective mVSS classification was used as the reference standard and analyses were restricted to the data where there was absolute agreement between R1 and R3.

A sample size of 100 subjects was recruited to ensure a minimum of 20 subjects in each of the mVSS categories of pigmentation and vascularity to conduct a meaningful exploratory analysis [16]. This sample size was also considered adequate to conduct Receiver Operating Characteristic (ROC) analysis, if feasible, with level of significance 0.05 and 80% power to detect a minimum ROC area under curve of 0.75 [16]. After recruitment it was decided that the mVSS classification category of mixed pigmentation would be excluded from the analysis due to the heterogeneous nature of mixed pigmented scars. A scar classified as mixed pigmentation by mVSS classification may include areas of hypopigmentation, normal pigmentation and hyperpigmentation and requires a different scar sampling protocol to that used in this study. There were insufficient data to perform ROC analyses. All data analyses were performed using Stata 12 statistical software (StataCorp LP, College Station, TX).

4. Results

4.1. Subject stratification and descriptive statistics

Of the 100 subjects recruited, 61% were male and the median age was 35 years (IQR: 26–52 years; minimum–maximum: 18–73 years). The anatomical locations of the assessed scars were: arm (n = 44), leg (n = 45), chest and abdomen (n = 6), head and neck (n = 1), groin and buttock (n = 1) and back (n = 3). The Fitzpatrick skin type (identified from outpatient medical records for 85 out of 100 subjects) for the subjects were: type II (n = 33), type III (n = 41), type IV (n = 7) and type V (n = 4). Final figures of stratification at recruitment are presented in Table 1.

4.2. The mVSS rater agreement

There was no disagreement in mVSS scar classification between R1 and R3 for 27/100 study subjects for pigmentation, and 34/100 study subjects for vascularity. Table 1 presents mVSS pigmentation and vascularity classifications for the total number of subjects recruited and where there was absolute agreement between R1 and R3.

4.3. Exploratory data analysis – pigmentation

Results of univariate descriptive statistics for the melanin index percent (M1%) for mVSS categories of pigmentation (normal pigmentation, hypopigmentation and hyperpigmentation) are presented in Table 2. The Kruskal–Wallis test with post-test comparisons indicated statistically significant differences in distribution between the following pigmentation categories ($\chi^2 = 16.84; p = 0.004; degrees of freedom: 3; n = 54$): hypopigmentation vs. normal pigmentation and hypopigmentation vs. hyperpigmentation; no statistically significant differences in M1% for normal pigmentation vs. hyperpigmentation were found. Kendall tau-b ($\tau_b$) correlation analyses found statistically significant positive associations of moderate
Table 1 – The number of subjects recruited using the mVSS (R1) and absolute agreement between R1 and R3 used in the analyses.

<table>
<thead>
<tr>
<th>mVSS category</th>
<th>Subjects recruited using mVSS (R1)</th>
<th>Absolute agreement between mVSS raters (R1 and R3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pigmentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>27</td>
<td>19</td>
</tr>
<tr>
<td>Hypopigmentation</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Hyperpigmentation</td>
<td>29</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>54</td>
</tr>
<tr>
<td>Vascularity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>Pink</td>
<td>28</td>
<td>16</td>
</tr>
<tr>
<td>Red</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td>Purple</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>66</td>
</tr>
</tbody>
</table>

* Mixed pigmentation category was excluded from analysis (27 subjects; 8 disagreements R1 vs. R3).

strength between the DermaLab Combo® MI% values for pigmentation and R1 mVSS (r = 0.4, p < 0.001) and R3 mVSS (r = 0.4, p < 0.001) pigmentation scores.

Three outliers of the normal pigmentation mVSS category were identified within the hyperpigmentation category and excluded from the analyses. These subjects had high vascularity readings by both the mVSS and the DermaLab Combo® (EI), and scars may have been misclassified as normally-pigmented instead of hyperpigmented. In the presence of a high vascularity score it is difficult to assess the “brownness” of the scar.

Dot plots for MI% scores for mVSS pigmentation categories, normal, hypopigmentation and hyperpigmentation are presented in Fig. 3. Half of the MI% values for scars classified as hypopigmented (mVSS) demonstrated a decrease in the MI% value relative to normal pigmentation (i.e. <100%). For scars classified as normal pigmentation by the mVSS, 8/19 had an MI% value close to normal (MI% 100–10%) with almost half of the MI% values demonstrating an increase in pigmentation of 10% or greater relative to normal (MI% >110%); for two cases, results indicated a decrease in pigmentation of greater than 10%. For scars classified as hyperpigmented by the mVSS, the majority of the MI% values showed a greater than 10% increase in pigmentation relative to normal (MI% >110%). However, for two cases MI% values were either close to normal (MI% 100–10%) or showed a decrease in pigmentation of greater than 10% (MI% <90%).

The mVSS classification of the DermaLab Combo® pigmentation results using univariate and bivariate approaches are shown in Fig. 4 and concordance results are summarised in Table 3. Bivariate plots of the MI% and EI% values by mVSS pigmentation categories are presented in Fig. 4. Fig. 4A shows univariate pigmentation classification (MI%) using cut-off values identified as 90%: hypopigmentation: 90–125%; normal pigmentation; and >125%: hyperpigmentation. Classification concordance was 60% (Table 3A) and increased by 2% with omission of three outliers from the analysis (Table 3B). Fig. 4B presents bivariate classification of pigmentation using discriminant cut-off values (p = 208: hypopigmented; 208–257: normal; >257: hyperpigmented). Using these discriminant cut-off values for pigmentation, 74% concordance was achieved (Table 3C); concordance was increased by 6% (80%) with omission of three outliers from analysis (Table 3D).

Classification of the data by mVSS pigmentation categories using two predictor variables (MI% and EI%) rather than one variable (MI%) resulted in improvement in classification concordance of 13% and 17%, for analyses using all values and values omitting outliers, respectively (Table 3).

4.4. Exploratory data analysis – vascularity

Results of descriptive statistics for the variable erythema index percent (EI%) for mVSS categories of vascularity (normal vascularity, pink, red and purple) are presented in Table 4. Kruskal-Wallis test with post-test comparisons identified statistically significant differences in EI% medians for vascularity categories of normal vs. pink, normal vs. red and normal vs. purple; no other pairwise comparisons of vascularity...
demonstrated statistically significant differences ($\chi^2$, degrees of freedom: 3, n = 66) = Chi square: 17.56 (p < 0.001). Kendall $\tau_b$ (\(\tau_b\)) correlation analyses found statistically significant positive associations of moderate strengths for DermaLab Combo $EI\%$ values for vascularity with $R1$ mVSS ($\rho_b$ 0.4, \(p < 0.001\)) and $R3$ mVSS ($\rho_b$ 0.3, \(p < 0.001\)) vascularity scores. Fig. 5 represents the dot plot of $EI\%$ values for vascularity categories of normal vascularity, pink, red and purple. For scars classified as normal vascularity by the mVSS, one case had an $EI\%$ value close to normal ($EI\%$ 100 ± 10%), 15/22 of $EI\%$ values showed a greater than 10% decrease in vascularity relative to normal ($EI\%$ < 90%) and 6/22 of $EI\%$ values showed a decrease between 10% and 20% relative to normal levels.

Fig. 4 – MI\% and EI\% values by mVSS pigmentation category. (A) Cut-off values on the x axis (MI\%); (B) cut-off values on the transformed axes.
Table 3 - Classification concordance results for M1% by mVSS pigmentation categories for cut-off values using standard and transformed x axis.

<table>
<thead>
<tr>
<th>mVSS</th>
<th>MI%</th>
<th>Hypo &lt;90</th>
<th>Normal 100–125</th>
<th>Hyper &gt;125</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cut-off values with x axis (MI%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. All scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypo</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>1</td>
<td>8</td>
<td>5</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Hyper</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>17</td>
<td>14</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>B. Without outliers*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypo</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Hyper</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>16</td>
<td>12</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>mVSS</td>
<td>MI%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cut-off values with transformed axis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. All scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypo</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Hyper</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>9</td>
<td>15</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>D. Without outliers*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypo</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Hyper</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>9</td>
<td>12</td>
<td>35</td>
<td></td>
</tr>
</tbody>
</table>

* Deleted outliers = (124,300) (135,275) (175,410).

greater than 10% increase in vascularity relative to normal (EI% > 110%). For scars classified as pink by the mVSS, nearly half of the EI% values (7/16) showed a greater than 10% increase in vascularity relative to normal (EI% > 110%). However, 6/16 had an EI% values close to normal (EI% 100 ± 10%) and in a small number of scars (3/16) the EI% values indicated a greater than 10% decrease in vascularity relative to normal (EI% < 90%). For scars classified as red by the mVSS, 12/16 of EI% values showed a greater than 10% increase in vascularity relative to normal (EI% > 110%). The remaining EI% values (4/16) were close to normal (EI% 100 ± 10%). The highest EI% value for scars classified as pink by the mVSS was 200%.

Comparing the data in Fig. 5, there is considerable overlap of EI% value between scars classified as pink and red using the mVSS. For 8/12 scars classified as purple by the mVSS had an EI% value with a greater than 10% increase in vascularity relative to normal (EI% > 110%). Three of the scars classified as purple by the mVSS had EI% values close to normal (EI% 100 ± 10%) and one scar had an EI% value with greater than 10% decrease in vascularity relative to normal (EI% < 90%). Comparison of the distributions of EI% values showed that 11/12 scars classified as purple by the mVSS had EI% values nested within the range of EI% values for scars classified as red.

Bivariate plots of the classification of EI% and M1% values by mVSS vascularity categories is presented in Fig. 6. Visual examination identified no visible clustering of data (univariate (EI%) or bivariate) by mVSS vascularity categories (normal, pink, red and purple) and no further analyses were conducted.

5. Discussion

This study has provided a useful starting point and valuable preliminary information in the use of DermaLab Combo® pigmentation and vascularity measurements in burns scar assessment. Our study showed that quantifying the relative percentage changes in erythema (EI%) and melanin (M1%) was a useful way to explore how the DermaLab Combo® pigmentation and vascularity measurements related to mVSS classification. In addition, positive associations of moderate strength were demonstrated between the mVSS clinical classification of pigmentation and vascularity and the DermaLab Combo® derived M1% and EI% values, respectively. However, further investigation is required to obtain evidence to support validity of the DermaLab Combo® pigmentation and vascularity measurements in burns scar assessment.

The results of this study showed that M1% values as a measure of scar pigmentation did not clearly fit within distinct ranges according to the mVSS pigmentation categories. Less than half of the scores classified as hypopigmentation (43%) and normal pigmentation (42%) using the mVSS had a M1% value that fitted within the appropriate range (M1% < 90, M1% 100 ± 10 respectively). However, for the majority (89.5%) of scars classified as hyperpigmented by the mVSS, M1% values were greater than 110%. The concordance associated with classification of mVSS pigmentation categories using univariate cut-off values of M1% was 63% (Table 7). Examination of mVSS pigmentation categorisation using the two variables, M1% and EI%, rather than M1% alone, improved pigmentation categorisation with greater concordance (80%). Changes in the skin chromophore concentrations (melanin and haemoglobin) induce changes in both MI and EI indices in narrow-band reflectance spectrophotometry, making it difficult to separate the contributions of each component [17] and this may explain
the improved concordance of categorisation using the two
variables.

The poor fit between the MI% values and the mVSS
categories of normal pigmentation and hypopigmentation
may be attributed to several factors. The low inter-rater
reliability of the pigmentation component of the mVSS means
that misclassification of mVSS categories was possible even
when the two raters agreed. In an earlier study we found that
inter-rater reliability for pigmentation in the mVSS was
dependent on the severity of the scar and the combination
of raters [7]. Another possible factor that may have contributed
to the poor fit between the MI% values and the mVSS
categories of normal pigmentation and hypopigmentation is
the influence of body hair on the melanin estimates obtained
with spectrophotometry [18]. While the index scar areas of our
subjects were hairless, the normal skin areas of our subjects
were not shaved, and this may have influenced the MI% scores.

The results of the study showed that EI% values did not fit
within ranges that reflected the mVSS vascularity categories.
The majority of scars (68%) classified as normal (mVSS) had an
EI% values less than normal (EI% < 90) and 27% of scars had
EI% values greater than normal (EI% > 110). There was
considerable overlap between the pink and normal categories.
Only 18.7% of red scars had an EI% values greater than 200%,
the highest EI% score in the pink category, and over 80% of red
scars were nested within the mVSS pink category. Approxim-
ately 92% of scars classified as purple had EI% values nested
within the mVSS red category. Examination of mVSS
vasculatity categorisation using two predictor variables, EI%

![Fig. 5 – Dot plot of EI% values of scars for vascularity by mVSS with summary statistics (median, minimum–maximum).](image)

![Fig. 6 – EI% and MI% values by mVSS vascularity category.](image)
and MI%, did not identify any patterns of clustering of values by mVSS (Fig. 6).

The poor fit between the DermaLab Combo® measurements and the mVSS vascularity categories may reflect the wide range in the degree of agreement between raters (weighted Kappa 0.4-0.8) reported for the vascularity component of the mVSS [7]. Another factor for consideration is that measurement of erythema by narrow-band reflectance spectrophotometry is very sensitive and may detect changes that are not solely due to scar vasularity, and as such, any changes in skin or scar erythema due to touch or palpation may be detected [19].

While this study provided an initial examination of the use of an objective scar assessment tool, the study had several limitations. The major limitation was the lack of an available gold standard in scar assessment and the use of subjective clinical expert opinion as the reference standard (mVSS) to compare DermaLab Combo® measurements. Expert opinion is considered to be a reasonable proxy when validating health instruments in the absence of a gold standard [20–22]. Future validation studies of the DermaLab Combo® measurements of pigmentation and vascularity may be possible using standard colour reference cards developed by the cosmetic industry, for example the Skin Colour Chart® (L’Oreál) [23]. The sample size for this study was based on sufficient baseline numbers within each mVSS category for both pigmentation and vascularity; however, this did not allow for disagreement between raters (R1 and R3), that resulted in a reduced sample size for analysis. The heterogeneous nature of the study sample may also have introduced a limitation. The inclusion of subjects from a range of skin types, burn scars at different stages of healing (data not shown) and scars in different anatomical locations created heterogeneity within the study sample. Despite the study sample being representative of our general patient population at RPH, it may have confounded our results. Campbell and Machin [9] discuss that a diagnostic assessment may have variable levels of sensitivity at different stages of a condition, and this may be applicable to our study. For future studies it is suggested that larger sample sizes be used where study subjects are stratified by skin type, stage of healing and the location of scar, to obtain more uniform samples for analyses. Study protocols that include a longer washout period (minimum of 20 min) between scar assessments and the shaving of normal skin areas prior to scar assessment are also recommended. Given the lack of a current gold standard measurement tool for scar assessment, further validity studies using the mVSS as standard reference may be improved using a greater number of raters to improve the accuracy of the assessment.

The mVSS mixed pigmentation classification refers to a scar with uneven areas of hypo-pigmentation, normal pigmentation and hyperpigmentation. A measure of spread of the DermaLab Combo® pigmentation measurement will vary depending on the heterogeneity of the scar pigmentation composition and also the skin type of the individual. To progress the understanding of the DermaLab Combo® measurements in context of the clinical category of mixed pigmentation (mVSS), future studies may consider a protocol that either identifies specific sub-sites of hypo, normal and hyper pigmentation within the heterogeneous pigmented scar for data collection and ongoing surveillance or that systematically identifies a greater number of sub-sites (e.g. grid) within the mixed pigmentation scar.

The major strength of this study is that it has provided useful preliminary information in the use of DermaLab Combo® pigmentation and vascularity measurements in burns scar assessment. Our study showed that quantifying the relative percentage changes in erythema and melanin, that is the increase or decrease in values of scar compared to the normal skin, was a useful starting point in understanding the DermaLab Combo® pigmentation and vascularity measurements. The results for pigmentation using MI% showed clustering of scores into mVSS categories was further improved by incorporating EI%. The results for vascularity using EI%, however, did not show any such improvements of clustering of scores into mVSS categories by incorporating MI%. This study also provides a useful research platform from which further investigations can be undertaken with respect to strengthening protocols, including sampling strategies, and the need to pursue other means or tools to assess the validity of the DermaLab Combo® measurements that will have clinical utility in burn scar assessment.

6. Conclusion

This study provides initial evidence in the interpretation of the DermaLab Combo® measurements in the clinical context of burns scar assessment. Results of our study showed that MI% values derived from the DermaLab Combo® measurements could be used to differentiate pigmentation categories classified by the mVSS (hypopigmentation, normal pigmentation and hyper pigmentation). The DermaLab Combo® is an objective tool providing continuous numerical data that may be useful for clinical scar monitoring over time; however, further work is required to understand and optimise the interpretation of the DermaLab Combo® measurements for pigmentation and vascularity. Currently, scar assessment at RPH Burns Unit and other burns care facilities use the subjective VSS or a modified version (i.e. mVSS). Therefore, improvements in the interpretation of the DermaLab Combo® measurements and classification of burn scar parameters will be valuable.

Conflict of interest statement

All authors declare no conflict of interest.

Acknowledgements

The authors acknowledge the funding support of the Wound Q5 Management Innovation Cooperative Research Centre. The authors also acknowledge Anne Henderson and Sarah McGarry for their contribution to the study as scar raters. The authors thank Maggie Crowe for her help with recruitment and Sugeesh Ariyaratna for his contribution in proof reading of this paper.
Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.burns.2015.01.012.

REFERENCES
