A study examining the reliability of digital ulcer definitions as proposed by the UK Scleroderma Study Group: challenges and insights for future clinical trial design

Hughes M[1], Bhushan M[2], Chakravarty K[3], Denton C[3], Dubey S[4]; Guiducci S[5]; Muir L [6]; Ong V[3]; Parker L[3]; Pauling J[7]; Prabu A[8]; Ravenscroft J[9]; Rogers C[10]; Roberts C[11]; Tracey A[1] and Herrick A[1]

[1]Centre for Musculoskeletal Research, The University of Manchester; [2]Department of Rheumatology, Blackpool Teaching Hospitals NHS Foundation Trust; [3]Centre for Rheumatology and Connective Tissue Diseases, Royal Free Hospital; [4]University Hospitals Coventry and Warwickshire NHS Trust,; [5]University of Florence; [6]Department of hand surgery, Salford Royal NHS Foundation Trust; [7]Department of Pharmacy and Pharmacology, University of Bath; [8]Rheumatology Department, City Hospital, Sandwell and West Birmingham Hospitals NHS Trust; [9]Member of Working Group, contact via Professor Herrick, The University of Manchester; [10]The University of Manchester; [11]Centre for Biostatistics, The University of Manchester.

Introduction

- The reliability of clinician grading of SSc-digital ulcers (DUs) has been reported to be poor to moderate at best [1-3], which has important implications for clinical trial design.
- To develop new DU definitions, a United Kingdom Systemic Sclerosis Study Group (UKSSG) working group was assembled.
- This comprised of 8 UK-based rheumatologists with an interest in SSc, an international SSc expert, a dermatologist, a hand surgeon and a rheumatology specialist nurse, 2 patients with SSc, and a specialist statistician.
- The proposed UKSSG definitions are presented in Table 1.

Aims

- The main aim of this study was to examine the reliability of new proposed UKSSG DU definitions ('no ulcer', 'healed ulcer' and 'DU') amongst UK clinicians with an interest in SSc.
- A secondary aim was to examine the performance of the definitions to be used in the context of 'preventative studies' (i.e. no ulcer vs healed ulcer/DU) and 'treatment studies' (i.e. no ulcer/healed ulcer vs DU).

Method (1)

- Raters graded through a custom-built web-based interface 90 (80 unique and 10 repeat) images of a range of digital lesions collected from patients with SSc (used in our previous study [3]).
- Lesions were graded on an ordinal scale of severity: 'no ulcer', 'healed ulcer', or 'DU'.
- No example images were given.

Method (2)

- Reliability was assessed using a weighted kappa coefficient, with bootstrapping to generate estimates of confidence intervals.
- Data was dichotomised by adjoining adjacent categories for application in future 'preventative' and 'treatment' studies for DUs.

Digital ulcer: A lesion (on the finger on or distal to the metacarpophalangeal joint) with loss of surface epithelisation and a visually discernible depth. The ulcer bed is often wet in appearance with surface slough. The peri-lesional skin surrounding digital ulcers is not uncommonly erythematous and/or macerated (including in the absence of superadded infection). Patients often report pain (which may be severe) associated with digital ulcers. Digital ulcers often have an overlying scab (eschar) and if there is a high index of suspicion of an underlying digital ulcer, then the lesion should be classified as such. Common sites for digital ulcers include the fingertips and over the extensor (dorsal) aspects of the hands, and in relation to subcutaneous calcinosis. Less often digital ulcers may occur at other sites on the hands (e.g. over the lateral aspects of the digits and at the base of the nail)**.**
Healed ulcer: A lesion with complete surface epithelisation (otherwise the lesion would be classified as a 'digital ulcer').
No ulcer: Any lesion which does not fulfil the definitions of either a 'digital ulcer' or 'healed ulcer' including (but not limited) to: digital pitting scars, hyperkeratosis, and fissures.

Table 1: UKSSG DU definitions

Results (1)

- 23 clinicians: 18 rheumatologists, 3 dermatologists, one hand surgeon and one specialist rheumatology nurse, completed the study.
- A total of 2070 (1840 unique + 230 repeat) image gradings were obtained.
- For intra-rater reliability, across all images the overall weighted kappa coefficient was high (0.71) and was moderate (0.55) when averaged across individual raters.

Results (2)

- Intra-rater reliability (kappa) was high for both the dichotomised analyses of 'preventative' (0.70, 95% CI = 0.62 – 0.79) and 'treatment' (0.77, 95% CI = 0.67 – 0.86) analysis. Inter-rater reliability was fair for the dichotomised analyses of 'preventative' (0.25, 95% CI = 0.19 – 0.31) and moderate for 'treatment' (0.41, 95% CI = 0.33 – 0.49) analysis.
- Figure 1 below illustrates a number of example images with high or low agreement between raters.

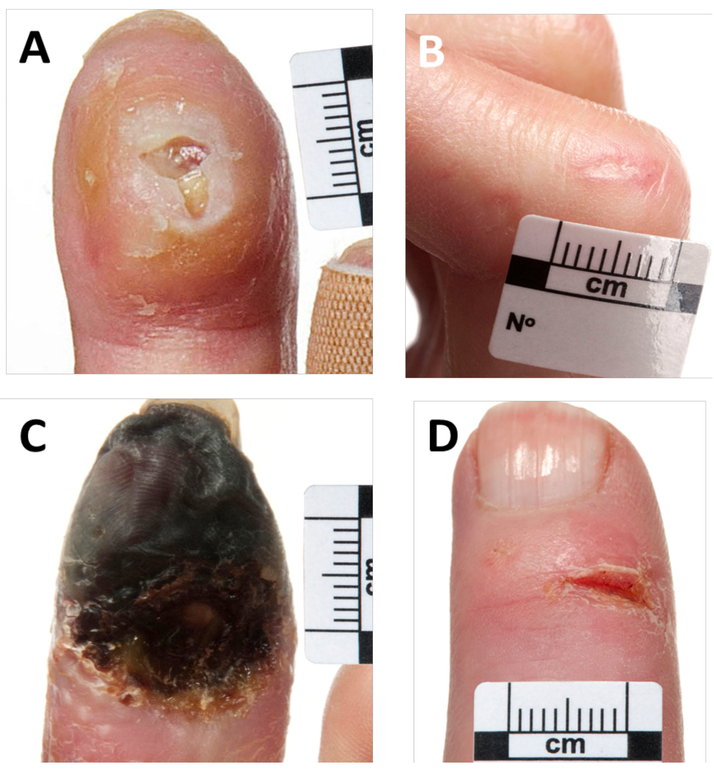

Figure 1: Example images of the proposed UKSSG DU definitions demonstrating different degrees of agreement among raters. A: High agreement (23 'DU'). B: High agreement (3 'no ulcer', 20 'healed ulcer', 0 'DU'). C: Low agreement (10 'no ulcer', 0 'healed ulcer', 13 'DU'). D: Low agreement (16 'no ulcer', 0 'healed ulcer', 7 'DU').

Conclusion

- Although our proposed DU definitions had high intra-rater reliability, the overall inter-rater reliability was poor.
- Our study further highlights the challenges of DU assessment by clinicians with an interest in SSc, and also provides a number of useful insights for the design of future clinical trials.
- Further research is warranted to improve the reliability of DU definition/rating as an outcome measure in clinical trials, including the role for objective measurement techniques and DU patient reported outcome measures.

1. Herrick et al 2009, Baron et al 2014, 3. Hughes et al 2016

Dr Michael Hughes/ Michael.hughes-6@postgrad.manchester.ac.uk