UNSW SYDNEY

Outcome measures for scoring the clinical signs of atopic dermatitis

Author:
Zhao, Cathy

Publication Date:
2016

DOI:
https://doi.org/10.26190/unsworks/19600

License:
https://creativecommons.org/licenses/by-nc-nd/3.0/au/
Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/57664 in https://unsworks.unsw.edu.au on 2024-05-03

Outcome Measures for Scoring the Clinical Signs of Atopic Dermatitis

A Thesis/A Dissertation

By

Dr Cathy Y. Zhao

Department of Dermatology, St George Clinical School

UNIVERSITY OF NEW SOUTH WALES

Submitted in partial fulfillment of the requirements

for the degree of

Master of Science (in Medicine)

Submitted on 31st of August, 2016

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: Zhao

First name: Cathy Other name/s: Yunjia

Abbreviation for degree as given in the University calendar: MSc

School: St George Clinical School Faculty: Medicine

Title: Outcome measures for measuring the clinical signs of atopic dermatitis

Abstract 350 words maximum: (PLEASE TYPE)

Background: Despite the existence of multiple outcome measures for the clinical signs of atopic dermatitis (AD), there is a need to standardise and validate them, especially for skin of colour patients.
Primary objectives: To compare the reliability of commonly used AD outcome measures and to validate them in skin of colour patients.
Methods: First study: 12 AD patients were scored by 5 clinicians using Eczema Area Severity Index (EASI), objective SCORing AD score (oSCORAD), Six Area Six Signs AD (SASSAD) and Three Items Severity (TIS). Second study: photos of 18 AD patients of various skin colours were scored by 5 clinicians using EASI, oSCORAD, SASSAD and TIS. Third study: 25 AD patients of various skin colours were scored by 5 clinicians over 2 scoring sessions using EASI, oSCORAD, IGA and greyscale. Melanin index was measured using a mexameter. Reliability was assessed using the intra-class correlation coefficient (ICC).
Results:
First study - Inter-rater reliability: the EASI and SASSAD showed good ICCs of 0.730(95%CI: 0.500-0.900) and 0.680(0.440-0.880), but TIS and oSCORAD had poor ICCs; Intra-rater reliability: EASI and TIS showed excellent ICCs of 0.886(0.744-0.952) and 0.820(0.614-0.923), while SASSAD showed a good ICC and oSCORAD had a poor ICC.
Second study had poor validity and was improved upon by the third study, where the inter-rater ICCs were:
EASI 0.827(0.658-0.941) in lighter patients and 0.774(0.598-0.906) in skin of colour; oSCORAD 0.680(0.441-0.880) in lighter patients and 0.736(0.544-0.889) in skin of colour; IGA 0.803(0.618-0.932) in lighter patients and 0.696 (0.490-0.868) in skin of colour; grey-scale had an ICC of 0.638(0.400-0.838) alone and 0.776(0.601-0.907) when replacing EASI's erythema scale. All scores had excellent intra-rater reliability. Erythema did not contribute to variability using coefficient of variance analysis.
Conclusions: EASI demonstrated excellent reliability in patients of all skin colours, supporting it as the core measure for AD clinical signs.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

FOR OFFICE USE ONLY Date of completion of requirements for Award:

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

COPYRIGHT STATEMENT

‘I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.
I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).
I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed …….....

Date …….....

AUTHENTICITY STATEMENT

‘I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.’

Signed …….....

Date …….....

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed ……..

Date ……..

Table of Contents

Dedication **4**

Acknowledgements **5**

Chapter 1. Background **6**
Standardising outcome measures for atopic dermatitis (AD) **6**
AD in skin of colour patients **8**
Outcome measures investigated in this thesis **10**
Harmonizing outcome measures for eczema (HOME) publications **29**

Chapter 2. Pilot comparison study of four AD severity scales **30**
Aims **30**
Method **30**
Sample size calculation 30
Participants and assessors 30
One-day scoring session 31
Statistical analysis 32
Results **33**
Patient demographics 33
Feasibility 33
Intra-rater reliability 34
Inter-rater reliability 37
Correlation between patient-reported and clinician-reported measures 40
Discussion **41**

Chapter 3. Virtual study of AD outcome measures for skin of colour patients **45**
Aim **45**
Method **45**
Participants and assessors 45
Scoring sessions 45
Categorisation of skin darkness 46
Statistical analysis 46
Additional assessments by an overseas dermatologist 47
Results **48**
Patients demographics 48
Intra-rater reliability 49
Inter-rater reliability 49
Erythema's contribution to variability 55
Convergent construct validity 56
Additional assessments by an overseas dermatologist 57
Discussion **57**

Chapter 4. Multi-centre study of AD outcome measures for skin of colour patients **60**
Aims **60**
Method **60**
Sample size calculations 60
Participants and assessors 60
Two independent half-day scoring sessions 61

Categorisation of skin darkness 62
Statistical analysis 63
Results 63
Patient demographics 63
Intra-rater reliability 64
Inter-rater reliability 65
Erythema's contribution to variability 66
Correlation between POEM and clinician-reported measures 66
Discussion 67

Chapter 5. Conclusions 69

Publications / related editorials /conference presentations associated with this thesis ...70

References 72

Dedication

To the pillars of my life: my husband Richard, and my parents Huiling and Hong, I would not have made it without your unconditional love and support, thank you.

Acknowledgements

I would first like to express my gratitude to my supervisor Professor Dedee Murrell for allowing me to be her Master of Science student. Thank you for steering me in the right direction and always being here for me when I run into problems. I am also appreciative of St George Clinical School of Medicine and the University of New South Wales for the tremendous amount of support I have received. The Australian Postgraduate Award and the Postgraduate Research Student Support scheme has helped me financially for completing and presenting this Master of Science at the World Congress of Dermatology, Vancouver. I would also like to thank the various co-investigators of the studies, especially Dr Michelle Rodrigues, Professor John Su, Dr Benjamin Daniel and Dr Linda Martin, who provided passionate input and participation to allow such smooth conduction of the studies. I am also grateful to Professor Matthew Law, who provided guidance with the sample size calculations and statistical methods used in this study. Furthermore, I would like to thank the staff and various junior doctors of Premier Specialists and Laser Dermatology, who provided help for the scoring sessions. Finally, and most importantly, I am indebted to the AD patients and their families for volunteering their time and participation in our studies.

Chapter 1. Background

Standardising outcome measures for atopic dermatitis (AD)

Atopic dermatitis (AD) is a chronic inflammatory skin condition affecting patients of all skin colours. According to various paediatric epidemiological studies, AD has a prevalence of 17% in United States, 14% in the United Kingdom, 24% in Japan, 17% in Korea, 17% in South Africa, 20% in Kenya and 32% in Australia,[1-7] with its prevalence shown to be gradually increasing in Africa, Eastern Asia and many parts of Europe.[8]

Currently, there are a myriad of AD interventional options being investigated in clinical trials. However, the evidence produced are difficult to be systematically compared due to the lack of fully validated and standardised AD outcome measures. This problem has prompted the Harmonizing Outcome Measures for Eczema (HOME) initiative, which consists of a collaborative group of international patients, clinicians, researchers and pharmaceutical industry representatives seeking to establish a core set of AD outcome measures.[9-11] In 2010, HOME met for the first time (HOME I) in Munich, Germany to introduce their goal and the need to involve carers and a multidisciplinary team in standardising AD outcome measures.[10] In 2011, HOME met again (Home II) in Amsterdam, Netherlands, and conducted a consensus voting exercise with 43 individuals (patients, clinicians, researchers, pharmaceutical industry) regarding the key domains of AD.[9] They concluded that there are four domains which should be included in all future AD trials as primary or secondary endpoints: clinical signs, symptoms, long-term control of flares and quality of life. Each of these domains will require a standardised core outcome measure.

In 2013, HOME met in the San Diego, United States regarding the core outcome measure for the clinical signs domain of AD (HOME III).[11] It was agreed that the intensity and extent of erythema, excoriations, oedema/papulation and lichenification should be included in the core outcome measures. A systematic review conducted as part of the HOME initiative which identified 16 different scales has been developed for the clinical signs of AD.[12-28] Five of the most commonly used ones include the objective SCORing Atopic Dermatitis (oSCORAD), Eczema Area and Severity Index (EASI), Six Areas, Six Sites Atopic Dermatitis (SASSAD), Three Item Severity index (TIS) and the Patient-oriented Eczema Measure (POEM).[22-27] Of these, only the oSCORAD and the EASI have been evaluated extensively. Although the oSCORAD has been shown to be a valid, internally consistent, and responsive score, its intra-rater reliability has not been evaluated.[23,29,30] Meanwhile, the EASI has been shown to be a reliable, internally consistent, valid and responsive measure.[24,30,31] It is a well-known measure with the advantage of giving separate severity scores to different body areas. When choosing between the oSCORAD and the EASI, as part of HOME III meeting, the HOME group conducted another voting consensus and agreed upon EASI to be the core outcome measure that should be used in all AD trials.[32] If desired, researchers can choose to use the oSCORAD *in addition to* the EASI.

Despite the multiple HOME meetings conducted to date, there has been no study directly comparing the EASI and oSCORAD head-to-head, especially regarding their reliability. One of the main objectives of this master thesis and the first study (Chapter 2) is to contribute to the growing process of AD outcome measures standardisation by directly comparing the reliability of the clinical signs measures oSCORAD, EASI, SASSAD and TIS (details of each measure will be explained later in this chapter).

The latest HOME meeting (Home IV) was in 2015 in Malmö, Sweden, which my supervisor Professor Murrell and I have attended and contributed.[33] During the meeting, a voting consensus was conducted to agree upon using the Patient Oriented Eczema Measure (POEM) as the core outcome for the symptoms domain of AD.

AD in skin of colour patients

Inflammatory skin disease such as AD are much harder to assess in patients with skin of colour, especially for clinicians unaccustomed to assessing skin of colour patients. For example, they may not appreciate the degree of inflammation if erythema is obscured by pigmentation giving greyish hues to the skin. Also, there are phenotypic variations in AD patients with dark skin, for example filaggrin-2 mutation variations have been more commonly found in African-Americans with AD, associated with a more persistent disease course.[34,35] Furthermore, environmental factors and cultural practices affect how the skin is cared for in darker skin patients, leading to further heterogeneity in AD clinical presentation. In this case, active AD may be mistaken for post-inflammatory hyperpigmentation.[36] Lastly, clinical experience with managing patients with skin of colour can vary considerably between clinicians, exacerbating the variability in assessment of AD in dark skinned patients.

After an extensive literature review from 1966 to present, we found assessing the clinical signs of AD in dark skinned patients is not well-recognised or well-addressed. There have been no previously validation studies addressing this issue besides a small study which showed that the skin type of AD participants in clinical trials conducted in the United States were under-reported, with over 40% of the trials from United States published between 2000 and 2009 omitting the AD participant's skin type.[37] Also, a systematic review revealed that

there is a dearth of clinical trials on efficacious systemic AD therapy in patients of racial and ethnic subsets in the United States.[38] Although, we do acknowledge that there is a difference between skin colour and ethnicity, as African American can vary significantly in skin colour.

Another key objective of this thesis is to address the issue of inaccurate assessment of clinical signs in dark skinned patients with AD. We conducted two studies altogether. The first was a pilot study conducted virtually in the full-body photographs of AD patients (Chapter 3). However, the validity of the study was rightly criticised when I orally presented it at the Skin of Colour Society Annual Meeting at the American Academy of Dermatology in Denver in 2015, as it utilised photographs instead of real-life AD patients.[39] Therefore, clinical signs including papulation, oedema and the heat of inflammatory could not be accurately assessed. Following this study, we reflected upon our limitations and re-conducted a multi-centre prospective study in real-life AD patients to compare the reliability of the EASI and oSCORAD (Chapter 4).

Outcome measures investigated in this thesis

Eczema Area Severity Index (EASI)

The EASI (Figure 1) is the recommended measure by HOME for the clinical signs of AD. It has formatting similar to the Psoriasis Area and Severity Index (PASI), wherein the extent of disease involvement and clinical signs (erythema, induration, excoriation and lichenification) are assessed to give a maximal composite score of 72.[31] The clinical signs are graded from 0-3, based on the average severity of the area. The extent of disease is determined through an estimation of the total body surface area (BSA) percentage in four areas: head and neck, trunk, upper and lower limbs.

Figure 1. The Eczema Area Severity Index (EASI) score.[31]

Figure 1. removed due to copyright reasons.

Objective SCORing Atopic Dermatitis (oSCORAD)

The SCORAD (Figure 2) focuses on three key areas: extent of disease intensity of clinical signs (erythema, oedema/papulation, crusting, excoriation and lichenification), and subjective symptoms (pruritus and sleep loss) to give a maximal composite score of 103. Intensity of clinical signs is graded from 0-3, based on a single representative area of the whole body.[23] The extent of disease is determined by the Rule of Nines. The oSCORAD was later proposed as a measure for the clinical signs of AD after removing the subjective symptoms of the SCORAD and bringing the maximal score from 103 to 83.[27]

Figure 2. SCORing Atopic Dermatitis (SCORAD).[23]

Figure 2 removed due to copyright reasons.

Six Areas, Six Sites Atopic Dermatitis (SASSAD)

SASSAD (Figure 3) is another score for measuring the clinical signs of AD. It assesses 6 signs of eczema (erythema, exudation, excoriation, dryness, cracking, and lichenification) on a scale of 0-3 in 6 areas (arms, hands, legs, feet, head and neck), hence yielding a maximal possible score of 108.[25] SASSAD removes the surface area component and eliminates the risk of over- or under-estimation of surface area involved.[40]

Figure 3. Six Areas, Six Sites Atopic Dermatitis (SASSAD).[25]

Figure 3 removed due to copyright reasons.

Three Items Severity Index (TIS)

TIS (Figure 4) was introduced in 1999 as a "simplified SCORAD" for assessing the clinical signs of AD. It contains an easy-to-administer 4-point scale (from 0-3) on the 3 signs (erythema, oedema and excoriation), with a maximum of 9 points.[22] The TIS is scored based on a representative site.

Figure 4. Three Items Severity Index (TIS).[22]

Figure 4 removed due to copyright reasons.

Investigator's Global Assessment (IGA)

The IGA is a type of outcome measure that score a global domain on a single ordinal scale. For our study, we have adapted the IGA from a previously published AD interventional study[41], which rates the clinical signs of AD on a 0-5 scale (Table 1).

Table 1: Investigator's global assessment used in this study.[41]

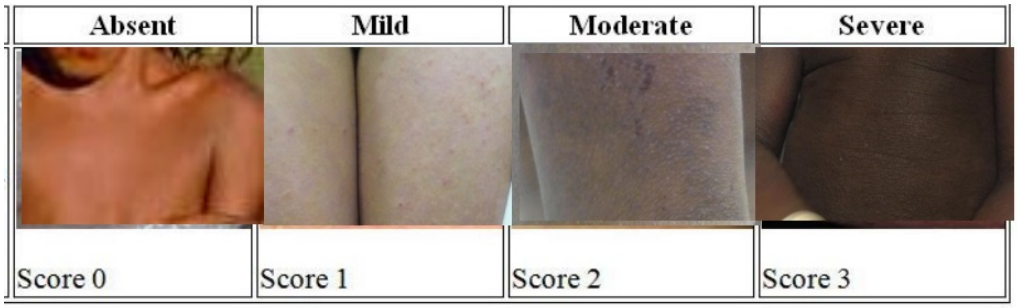
0 – Clear	No signs of AD
1 – Almost clear	Just perceptible erythema/papulation/infiltration
2 – Mild disease	Mild erythema/papulation/infiltration
3 – Moderate disease	Moderate erythema/papulation/infiltration
4 – Severe disease	Severe erythema/papulation/infiltration
5 – Very severe disease	Severe erythema/papulation/infiltration with oozing or crusting

Grey-scale

We adapted the grey-scale from the erythema component of the EASI (Figure 5). It only examines the clinical sign greyness, graded from 0-3, based on the average greyness of the area to give a maximal score of 18. The extent of disease is determined through the same method as the EASI. It was intended this could substitute the erythema score in EASI for skin of colour patients.

Figure 5. The greyscale used in the study for dark skinned patients, with examples illustrating its typical scores.

	Head/Neck	Upper Limbs	Trunk	Lower Limbs
Greyness (0-3)				
Area				



Patient-oriented eczema measure (POEM)

The POEM (Figure 6) assesses symptom frequency, the pattern of remission and relapse. The questionnaire consists of 7 questions, each with a 5-point scale (0-4) to give a maximal score of 28.[26] The questions involve the frequency of pruritus, sleep loss, bleeding, oozing, cracking, flaking and dryness. AD severity bands according to POEM scores have been proposed.[42] Recently, the HOME group recommended the POEM score as the core outcome measure for symptoms of AD in all future trials.[33]

Figure 5. Patient-oriented eczema measure (POEM).[26]

Figure 5 removed due to copyright reasons.

Dermatological Life Quality Index (DLQI)

The DLQI (Figure 7) is a measure of quality of life. It consists of 10 questions, scored on a scale from 0-3 to give a maximal score of 30.[43] It addresses impact on leisure activities, school / holidays, personal relationships, sleep and treatment. The Children's DLQI (CDQLI) (Figure 8) in a cartoon version was used for paediatric patients.

Figure 7. Dermatological Life Quality Index (DLQI)[43]

DERMATOLOGY LIFE QUALITY INDEX

DLQI

Hospital No: Date:
Name: Score:
Address: Diagnosis:

The aim of this questionnaire is to measure how much your skin problem has affected your life OVER THE LAST WEEK. Please tick one box for each question.

1.	Over the last week, how **itchy**, **sore**, **painful** or **stinging** has your skin been?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	
2.	Over the last week, how **embarrassed** or **self conscious** have you been because of your skin?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	
3.	Over the last week, how much has your skin interfered with you going **shopping** or looking after your **home** or **garden**?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	 Not relevant ❒
4.	Over the last week, how much has your skin influenced the **clothes** you wear?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	 Not relevant ❒
5.	Over the last week, how much has your skin affected any **social** or **leisure** activities?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	 Not relevant ❒
6.	Over the last week, how much has your skin made it difficult for you to do any **sport**?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	 Not relevant ❒
7.	Over the last week, has your skin prevented you from **working** or **studying**?	Yes No	❒ ❒	 Not relevant ❒
	If "No", over the last week how much has your skin been a problem at **work** or **studying**?	A lot A little Not at all	❒ ❒ ❒	
8.	Over the last week, how much has your skin created problems with your **partner** or any of your **close friends** or **relatives**?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	 Not relevant ❒
9.	Over the last week, how much has your skin caused any **sexual difficulties**?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	 Not relevant ❒
10.	Over the last week, how much of a problem has the **treatment** for your skin been, for example by making your home messy, or by taking up time?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	 Not relevant ❒

Please check you have answered EVERY question. Thank you.

©AY Finlay, GK Khan, April 1992 www.dermatology.org.uk, this must not be copied without the permission of the authors.

Figure 8. Children's Dermatological Life Quality Index (CDLQI).

Figure 8 removed due to copyright reasons.

Figure 8 removed due to copyright reasons.

SkinDex-29

The SkinDex-29 (Figure 9) is another measure of quality of life. It assesses 3 aspects of a patient's life: emotions, functioning, and symptoms. These are graded with 5-point scales giving a maximum score of 175.[44]

Figure 9. SkinDex-29 score.[44]

SkinDex-29

Name:

Date:

The aim of this questionnaire is to measure how much your skin condition has affected your life **IN THE LAST MONTH**. Please circle one response for each of the questions below. Young children should complete this questionnaire with the help of an adult.

1. My skin hurts

Never Rarely Sometimes Often All the Time

2. My skin condition affects how well I sleep

Never Rarely Sometimes Often All the Time

3. I worry that my skin condition may be serious

Never Rarely Sometimes Often All the Time

4. My skin condition makes it hard to work or do hobbies

Never Rarely Sometimes Often All the Time

5. My skin condition affects my social life

Never Rarely Sometimes Often All the Time

6. My skin condition makes me feel depressed

Never Rarely Sometimes Often All the Time

7. My skin condition burns or stings

Never Rarely Sometimes Often All the Time

8. I tend to stay at home because of my skin condition

Never Rarely Sometimes Often All the Time

9. I worry about getting scars from my skin condition

Never Rarely Sometimes Often All the Time

10. My skin itches

Never Rarely Sometimes Often All the Time

11. My skin condition affects how close I can be with those I love

Never Rarely Sometimes Often All the Time

12. I am ashamed of my skin condition

Never Rarely Sometimes Often All the Time

13. I worry that my skin condition may get worse

Never Rarely Sometimes Often All the Time

14. I tend to do things by myself because of my skin condition

Never Rarely Sometimes Often All the Time

15. I am angry about my skin condition

Never Rarely Sometimes Often All the Time

16. Water bothers my skin condition (bathing, washing hands)

Never Rarely Sometimes Often All the Time

17. My skin condition makes showing affection difficult

Never Rarely Sometimes Often All the Time

19. My skin is irritated

Never Rarely Sometimes Often All the Time

20. My skin condition affects my interactions with others

Never Rarely Sometimes Often All the Time

21. I am embarrassed by my skin condition

Never Rarely Sometimes Often All the Time

22. My skin condition is a problem for the people I love

Never Rarely Sometimes Often All the Time

23. I am frustrated by my skin condition

Never Rarely Sometimes Often All the Time

24. My skin is sensitive

Never Rarely Sometimes Often All the Time

25. My skin condition affects my desire to be with people

Never Rarely Sometimes Often All the Time

26. I am humiliated by my skin condition

Never Rarely Sometimes Often All the Time

27. My skin condition bleeds

Never Rarely Sometimes Often All the Time

28. I am annoyed by my skin condition

Never Rarely Sometimes Often All the Time

29. My skin condition interferes with my sex life

Never Rarely Sometimes Often All the Time

30. My skin condition makes me tired

Never Rarely Sometimes Often All the Time

Please check that you have answered EVERY question.

Thank you

Total Score =

Harmonizing outcome measures for eczema (HOME) publications

First meeting:

Schmitt J, Williams H. Harmonising Outcome Measures for Eczema (HOME). Report from the First International Consensus Meeting (HOME 1), 24 July 2010, Munich, Germany. The British journal of dermatology 2010;163:1166-8.

Second meeting:

Schmitt J, Spuls P, Boers M, et al. Towards global consensus on outcome measures for atopic eczema research: results of the HOME II meeting. Allergy 2012;67:1111-7.

Third meeting:

Chalmers JR, Schmitt J, Apfelbacher C, et al. Report from the Third International Consensus Meeting to Harmonise Core Outcome Measures for Atopic Eczema / Dermatitis Clinical Trials (HOME). The British journal of dermatology 2014.

Fourth meeting:

Chalmers JR, Simpson E, Apfelbacher CJ… Murrell, DF…et al, **Zhao CY**, Spuls P. Report from the fourth international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME initiative). Br J Dermatol 2016;175:69-79.

Chapter 2. Pilot comparison study of four AD severity scales

Aims

- To directly compare the feasibility, inter-rater and intra-rater reliability of the objective measures oSCORAD, EASI, SASSAD and TIS.
- Assess each score's correlation with patient-oriented measures, the DLQI/CDLQI, SkinDex-29 and POEM.

Method

This was a pre-planned prospective exploratory study. Ethical approval was granted by the South Eastern Sydney Local Health District HREC (Reference number: HREC/12/POWH/155). The scoring was conducted on 20th October, 2012.

Sample size calculation

According to the COSMIN guidelines, the sample size required for reliability analysis would be less than that of a factor analysis study.[45] Calculations showed that with a sample size of 11 patients and 5 assessors, an observed ICC of 0.9, 0.8 and 0.7 would be estimated with 95% confidence intervals of (0.81-0.99), (0.64-0.96) and (0.48-0.92) respectively. Thus the study would be powered to discriminate a high (0.9) from moderate (0.7) ICC, which we felt was an acceptable level of precision.

Participants and assessors

The AD patients were selected from the senior author's specialist dermatology practice based on a range of severities and ages, with the inclusion and exclusion criteria in Table 2.

Table 2. Inclusion and exclusion criteria for the study.

Inclusion criteria	Exclusion criteria
• Age range: 0-80 years • Those with AD, as per Hanifin and Rajka's criteria • Of mature mind and able to give consent, in case of children (below 18 years of age) then the guardian or caretaker can provide consent • Be able to understand and complete the DLQI/ CDLQI and Skindex-29 questionnaire (except for babies and infants)	• Participants who are unwilling or unable to comply with requirements of the protocol • Those withdrawing from the study. • Non-English speakers

All five assessors were either dermatologists or had completed full-time dermatology fellowships for two years. All five assessors have had long-term experience with scoring AD using the EASI and SCORAD. Just prior to the session, assessors also attended a one hour training session, where they reviewed all four scores and had their queries clarified.

One-day scoring session

On the scoring day, participants were placed in individual rooms asked to complete patient-oriented questionnaires. Each assessor followed a roster to rotate around the rooms. The assessors had 12 minutes to examine each patient using the 4 scales. Each scale was printed on a differently coloured sheet and these sheets were placed in random order for each assessor, per patient**.** At any one point, only one assessor was assigned to one patient. Each assessor recorded the time it took for them to complete each scale. To assess intra-rater reliability, each assessor was asked to rescore at least four randomly allocated patients. The randomisation process was kept undisclosed by an uninvolved investigator. Efforts were made to ensure at least half an hour time lapse between the assessors' first and second

attempt to minimise recall bias. The assessors were not allowed to discuss their scores with one another.

Statistical analysis

Statistical analyses were performed using Microsoft Excel and SPSS versions 18.0 and 22.0. Reliability was determined using one way random-effect ANOVA intra-class coefficient (ICC) ICC. When an ICC is below 0.40, the clinical correlation is poor; when it is between 0.40 to 0.59, the level of correlation is fair; when it is between 0.60 to 0.74, the level of correlation is good and when it is between 0.75 to 1, the level of clinical significance is excellent [46].

To determine body surface area (BSA)'s contribution to inter-rater variations, the EASI and oSCORAD had area scores graphed using scatterplot. The BSA component and the 'total minus BSA component' of the oSCORAD had its coefficient of variance (CV) calculated. The oSCORAD was chosen as only it has a directly extractable BSA component. The null hypothesis (that BSA does not contribute to oSCORAD inter-rater variations) is that the CV for the BSA component would be smaller than the 'total minus BSA component'.

A paired T-test was used to determine the significance of the difference. The correlation between clinician-oriented and patient-oriented scales were analysed using the two-tailed Spearman's rho and scatterplots.

Results

Patient demographics

Out of the 30 AD patients approached, 14 agreed to participate. On the day, 12 patients attended; 3 adults and 9 children. All patents were diagnosed with AD, with the exception of 1 patient also having ichthyosis vulgaris in addition. Table 3 shows the participants' characteristics. Four of the 5 assessors rescored 4 patients, and one assessor rescored 5.

Table 3. Patient demographics

Baseline characteristics		**AD Patients**
Number	Total	12
	Paediatric	9
	Adult	3
Age	Mean	13 (SD: ±6.4)
	Median	11
	Paediatric mean	9
	Adult mean	25
Sex (%)	Female	9 (75%)
	Male	3 (25%)
Ethnicity (%)	Caucasian	8 (67%)
	Asian	3 (25%)
	African	1 (8%)
Patient-oriented severity according to POEM* (%)	Moderate	6 (50%)
	Severe	5 (42%)
	Very severe	1 (8%)

*POEM score severity bands: 0–2 (clear/almost clear); 3–7 (mild); 8–16 (moderate); 17–24 (severe); 25–28 (very severe)

Feasibility

For feasibility, all scales were completed within 3 minutes. oSCORAD and SASSAD required on average 2:20 minutes, EASI required 2:10 minutes and TIS 40 seconds.

Intra-rater reliability

For intra-rater reliability, both EASI (ICC=0.886 [95% CI: 0.744-0.952] and TIS (ICC=0.820 [95% CI: 0.614-0.923]) demonstrated excellent reliability. SASSAD showed good reliability (ICC=0.720 [95% CI: 0.424-0.878]) whilst oSCORAD (ICC=0.446 [95% CI: 0.037-0.730]) had poor reliability. The score's Bland Altman plots are shown in Figure 10-13.

Figure 10. Bland-Altman plots demonstrating the variation in patient scores, which is a measure of the intra-rater reliability for the EASI.

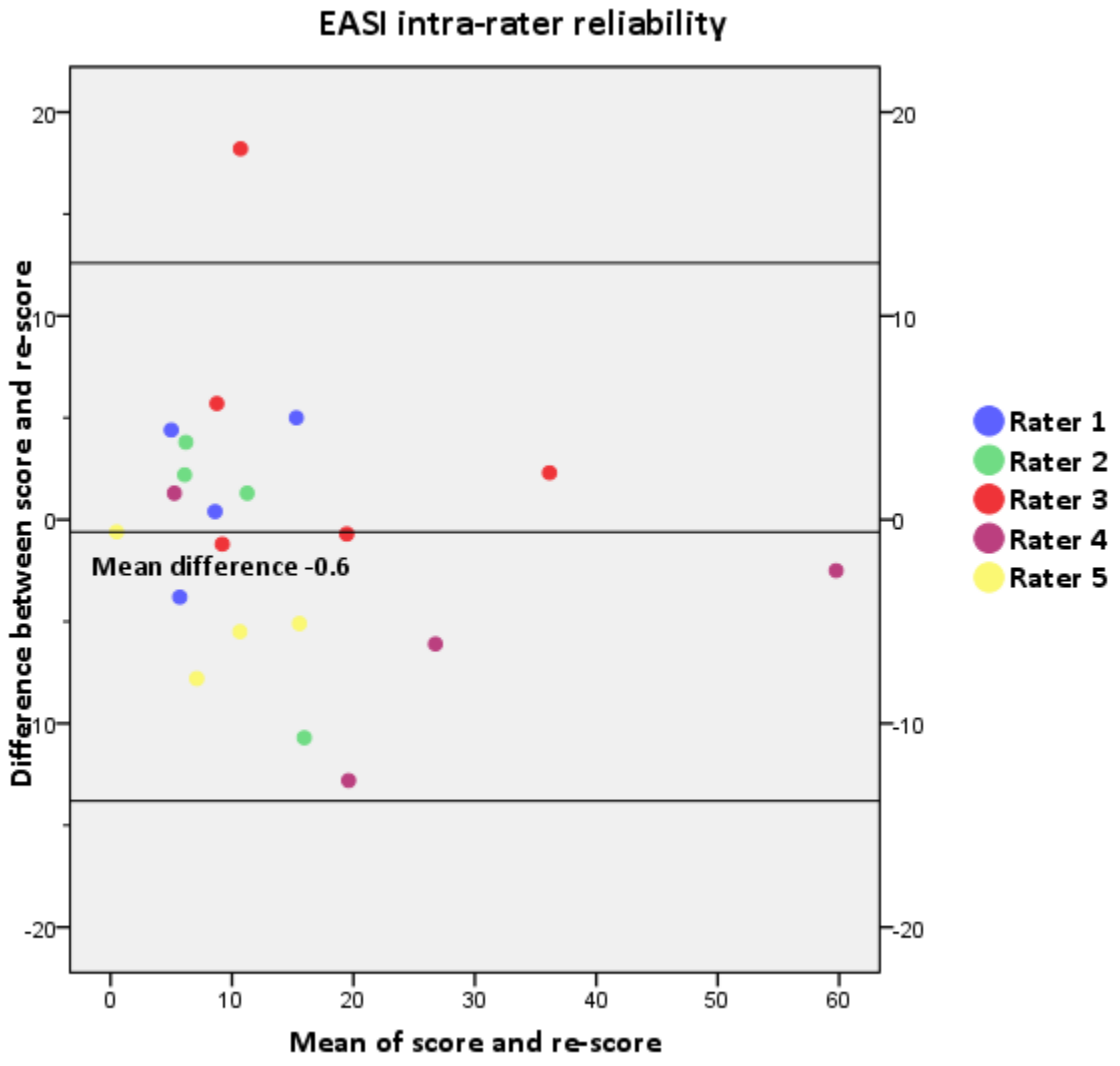


Figure 11. Bland-Altman plots demonstrating intra-rater reliability for the SASSAD.

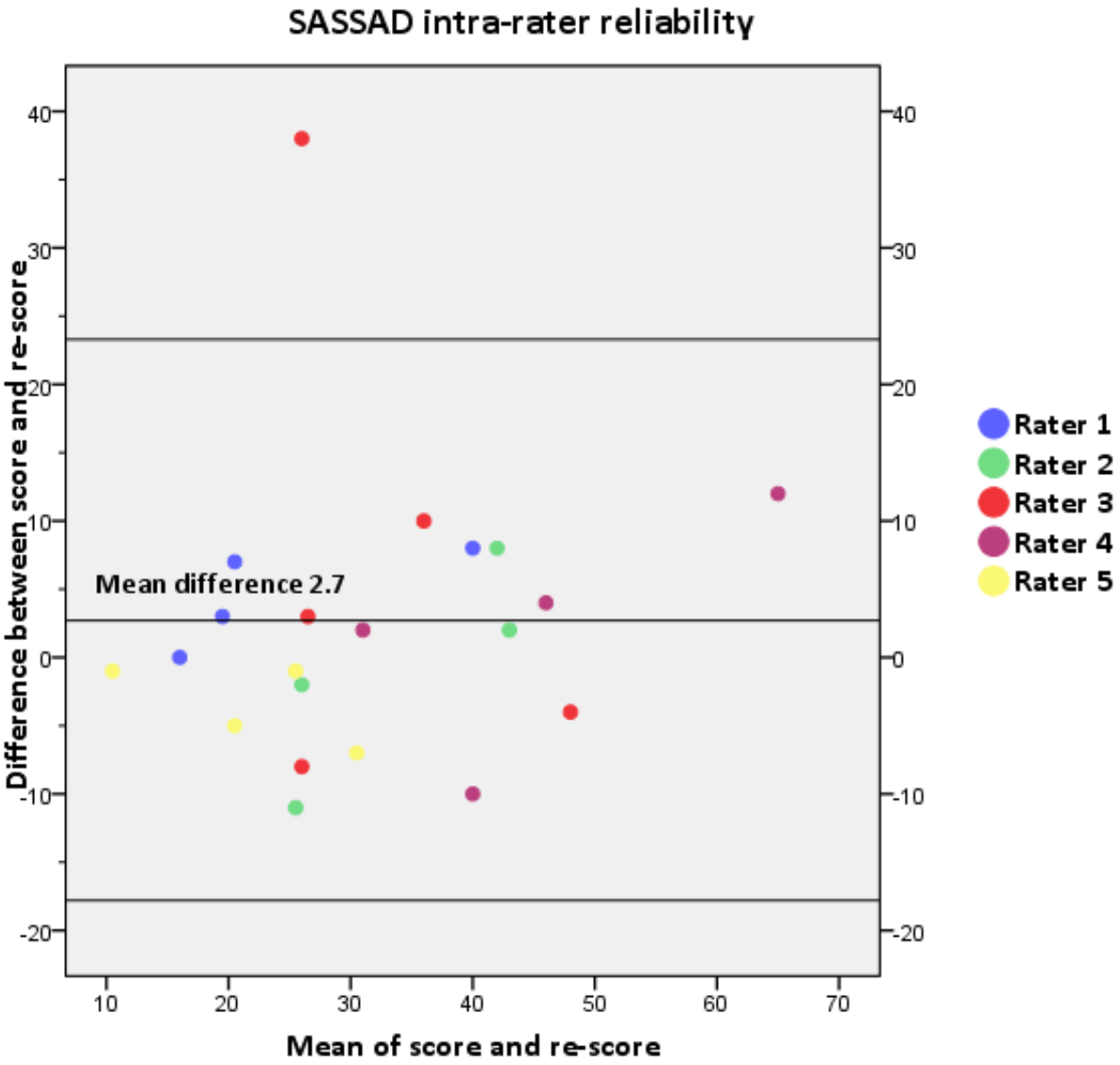


Figure 12. Bland-Altman plots demonstrating intra-rater reliability for the oSCORAD.

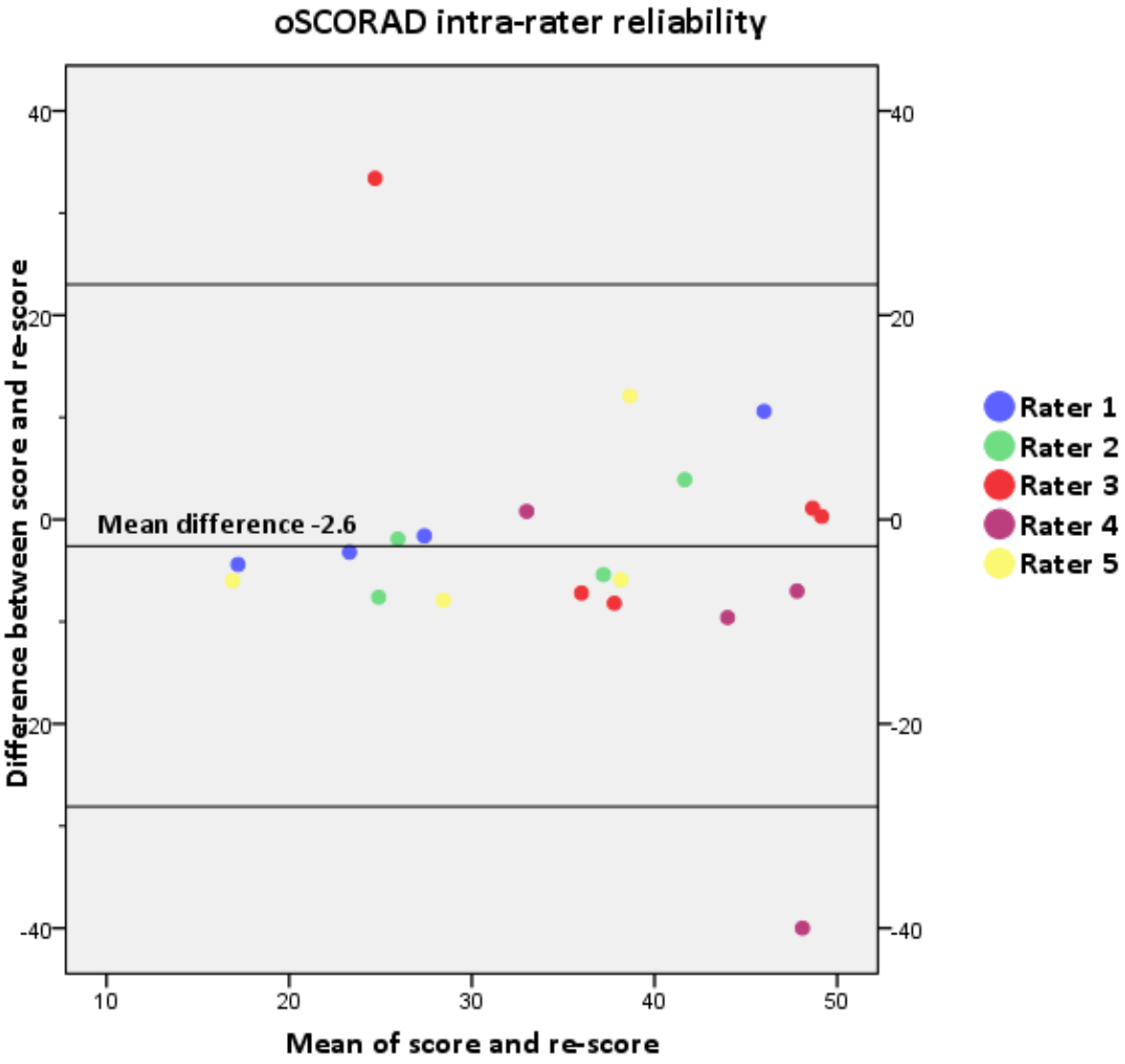
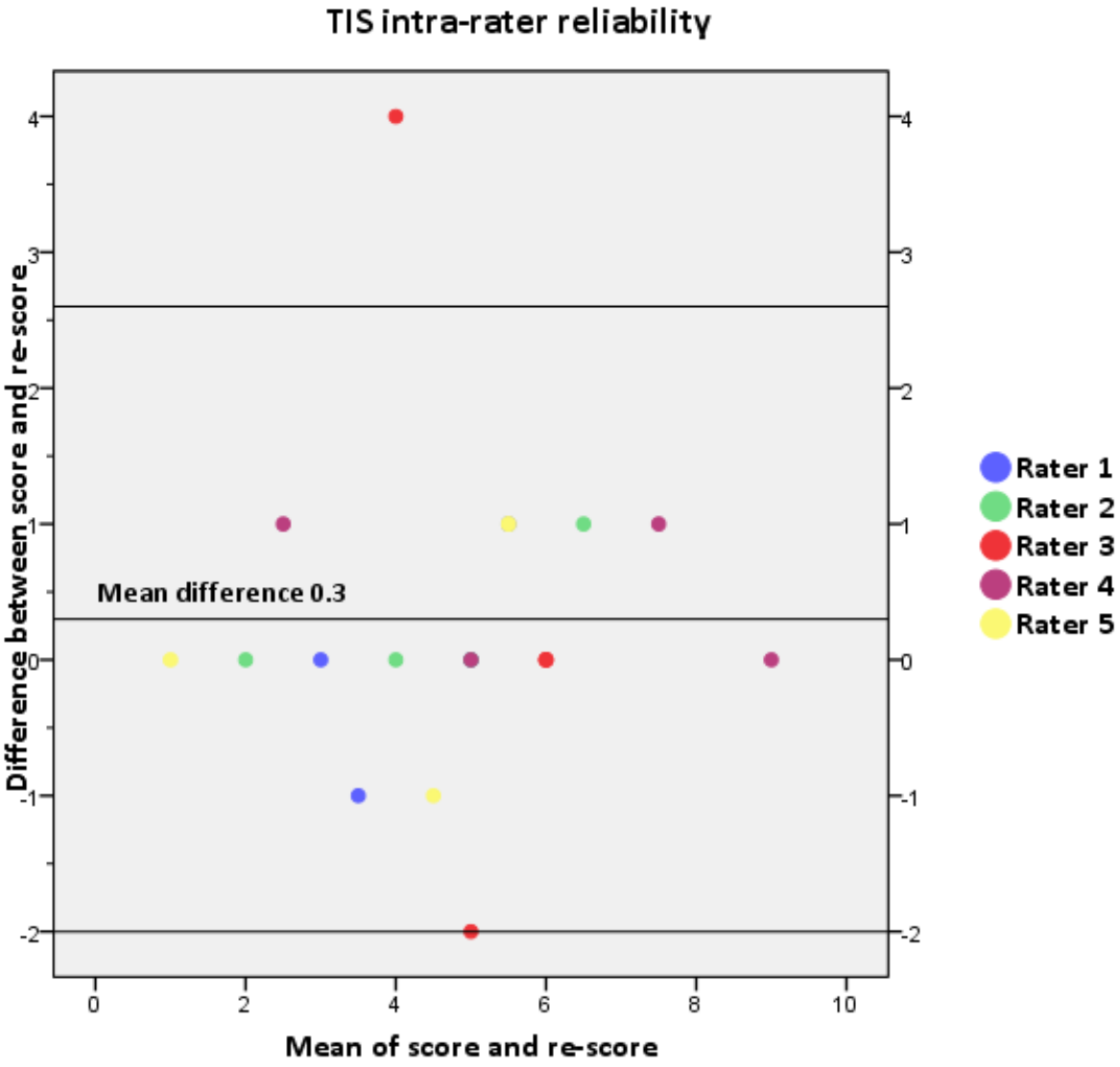


Figure 13. Bland-Altman plots demonstrating intra-rater reliability for the TIS.



Inter-rater reliability

For inter-rater reliability, both EASI (ICC=0.730 [95% CI: 0.500-0.900]) and SASSAD (ICC=0.680 [95% CI: 0.440-0.880]) demonstrated good reliability, whilst oSCORAD (0.498 [95% CI: 0.234-0.785]) and TIS (ICC=0.497 [95% CI: 0.233-0.785]) showed poor inter-rater reliability. Figure 14-17 demonstrated the score's inter-rater variations.

Figure 14. EASI's inter-rater scores in ascending patients' mean score rank.

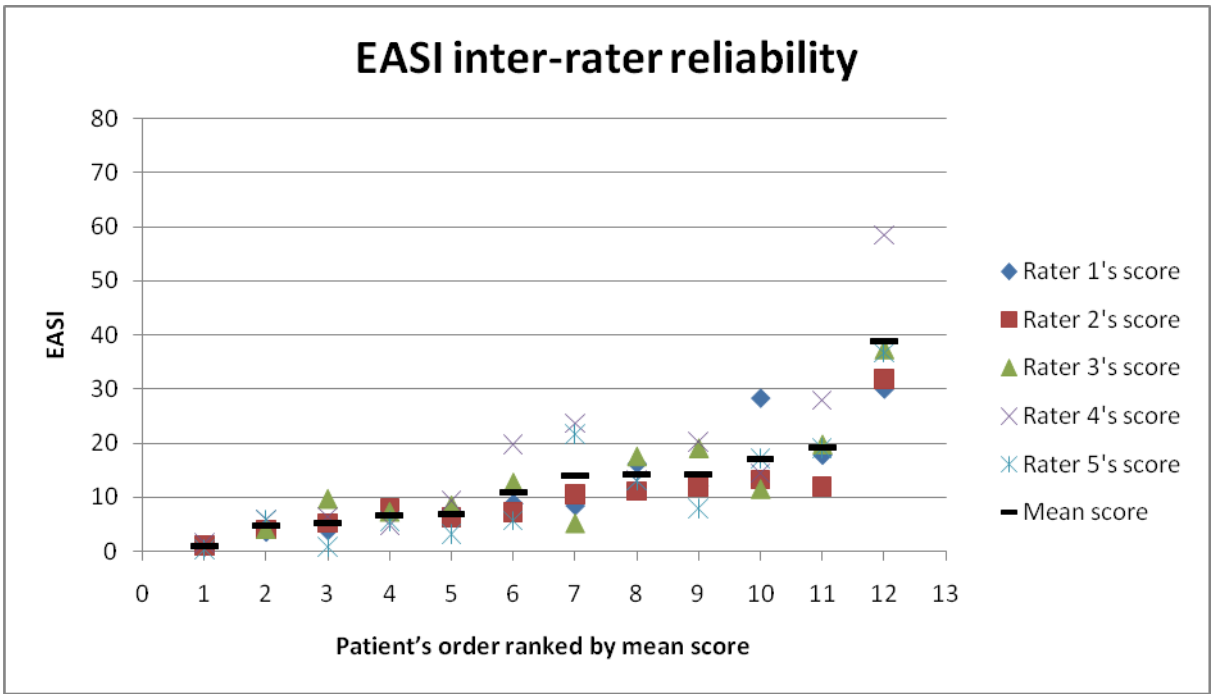


Figure 15. SASSAD'S inter-rater scores in ascending patients' mean score rank.

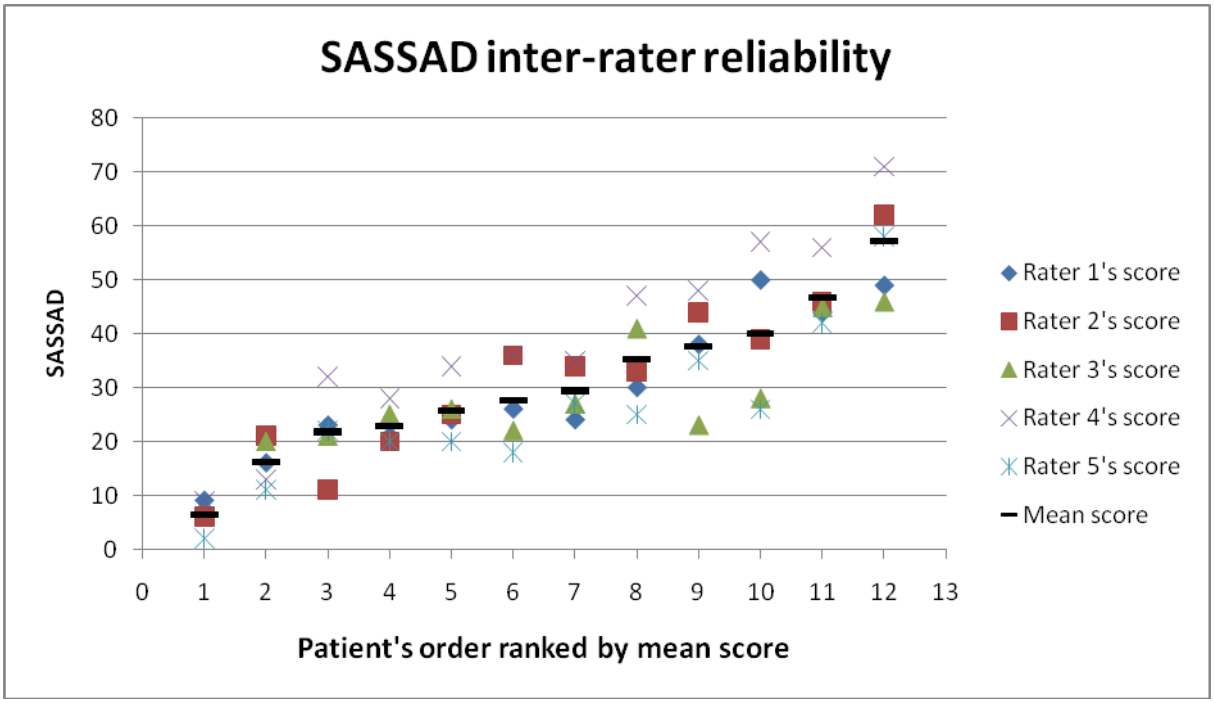


Figure 16. oSCORAD'S inter-rater scores in ascending patients' mean score rank.

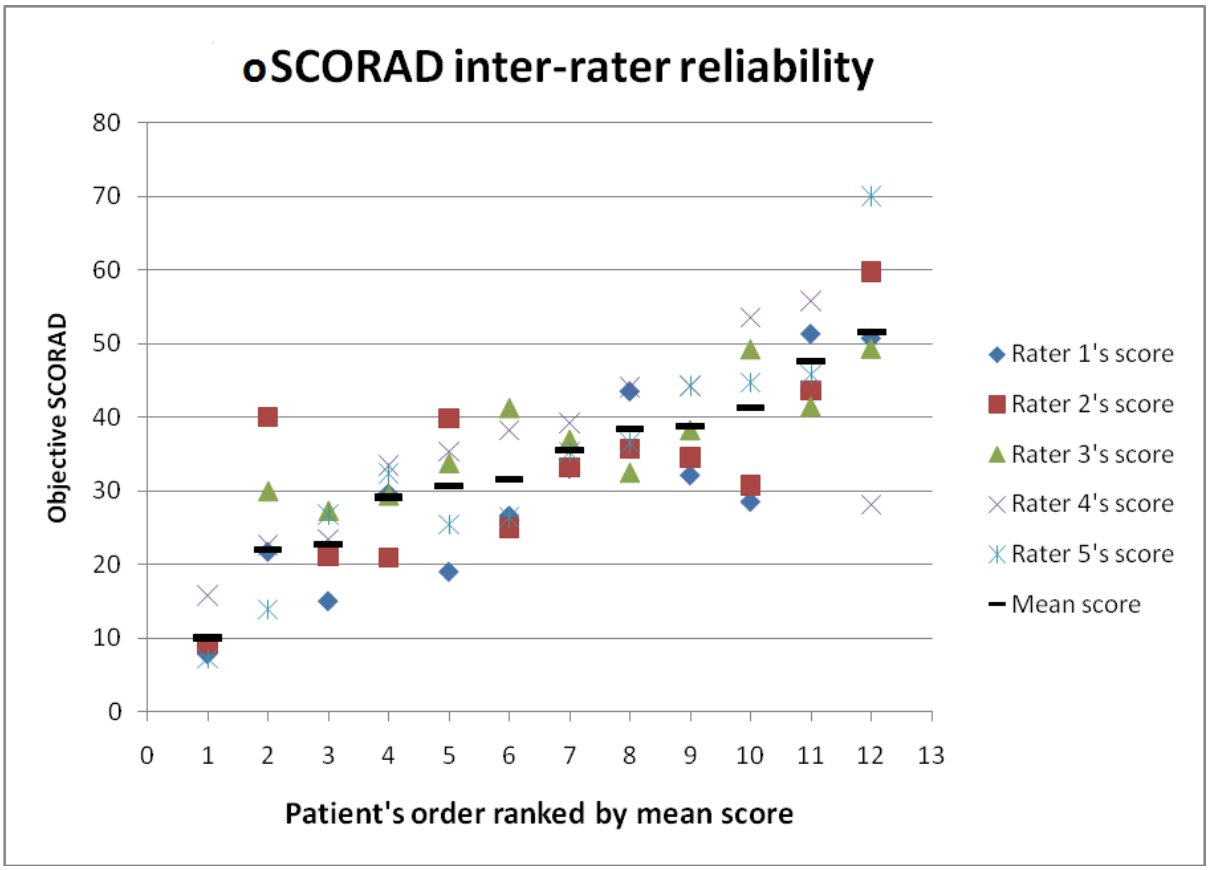
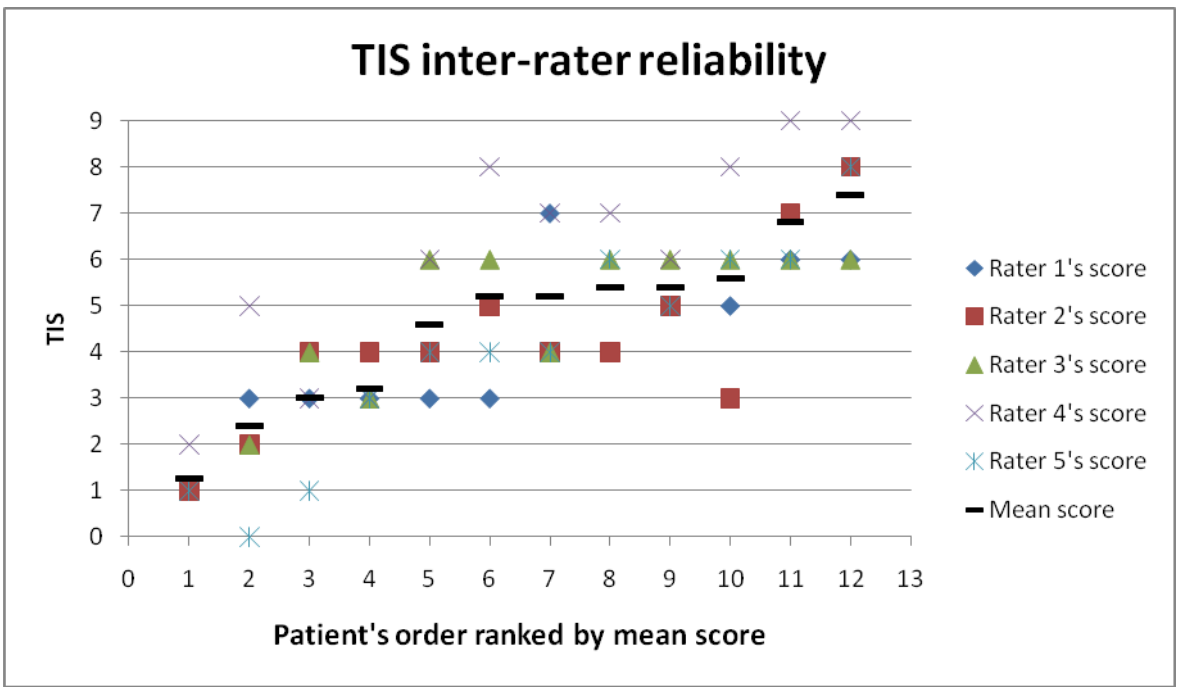


Figure 17. TIS' inter-rater scores in ascending patients' mean score rank.

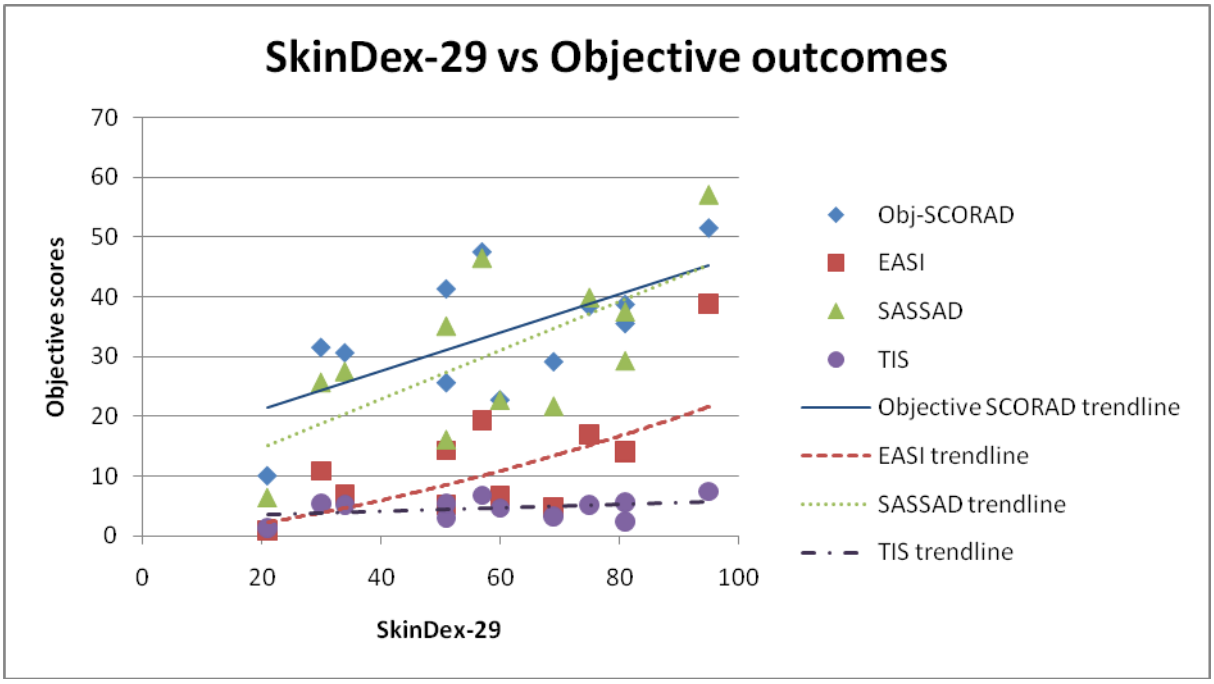


The inter-rater CV means of the oSCORAD BSA component was 0.404, higher than that of the 'total minus BSA component', which was 0.242. A paired T-Test showed a significant difference between these CV means with p=0.01. This showed that the BSA component of the oSCORAD contributed to its inter-rater variations.

Correlation between patient-reported and clinician-reported measures

Correlation between patient-reported and clinician-rated outcome measures showed moderate correlation between SASSAD and the SkinDex-29, ρ=0.611 (p=0.035) (Figure 18). The mean Skindex-29 was 58.6/120 (range: 21-95), mean DLQI/CDLQI score was 12/30 (range: 5-23) and mean POEM was 17/28 (range: 9-26). There was no correlation between the objective scoring systems and the DLQI/CDLQI and POEM.

Figure 18. Correlation between Skindex-29 and objective outcomes with trend lines.



Subjective feedback from the assessors showed that all assessors had preference of assessing AD severity in different sites, as in the EASI and SASSAD, instead of grading only a representative site. Also, all assessors felt that the TIS was fast and easy to use.

Discussion

This study showed that the EASI had consistently the most superior inter-rater and intra-rater reliability.[32] Interestingly, the SASSAD also showed moderate reliability, higher than the well-accustomed oSCORAD, questioning the suitability of the oSCORAD as a routine AD outcome measure.

The relatively poorer reliability of the well-accustomed oSCORAD in this study may be partially explained by its method of BSA estimation. BSA estimation has been agreed upon as an important aspect of AD severity assessment.[11] However, the method of BSA estimation varies between scores. The oSCORAD's poor inter-rater reliability may be a result of its utilisation of the palm size according to the "rule of nines" for an exact percentage, whereas EASI estimates the body's surface area using 7 choice bands. The complexity in separately scoring palm sizes may subject the assessor to further over- or underestimation of the affected area, especially in a child.[47] Despite our assessors being familiar with BSA estimation from clinical trials and the rule of nines, CV calculations have shown that the BSA did contribute to the inter-rater variations in oSCORAD scores. Meanwhile, an explanation for the SASSAD's relatively good inter-rater reliability is that it accounts for BSA by including 6 separate body regions, thus accounting for extent of involvement without having to estimate an exact area.

Disease severity in AD can vary according to the site affected. The generalisation of disease severity according to a representative site, independent of the body area affected may explain oSCORAD's poor inter-rater and intra-rater reliabilities as well as the TIS' poor intra-rater reliability. On the other hand, the EASI and SASSAD attempt to allow more accurate

representation of AD severity by having the different scores for all lesions at different sites. In particular, the SASSAD gives emphasis to the severity of patients' face, hands and feet in its more holistic patient assessment. These areas are usually more prominent in irritant and allergic contact dermatitis than in atopic dermatitis alone, but are particularly important when assessing impact of disease. Also, by thorough examination of disease severity of each independently affected body part, the most and least affected areas on the patients can be elucidated, thus allowing the prioritisation of targeted treatments such as UV therapy. In future validation studies, recording of the chosen 'representative sites' from SCORAD and TIS may elicit if certain areas are more suitable as 'representative sites'.

TIS' simplistic design allowed the fastest assessment time, excellent intra-rater reliability but surprisingly low inter-rater reliability. The low inter-rater reliability makes it questionable for use in multi-centre clinical trials and meta-analysis. However, TIS would benefit general practitioners as a screening tool with its good feasibility, as supported by subjective positive feedback from the assessors. However this may not apply to real clinical practice with increased amounts of distractions and interruptions.

Interestingly, we revealed low correlation between patients' perception of illness through quality of life questionnaires and dermatologists' objective assessment of disease severity. Among the quality of life instruments used in the study, only Skindex-29 showed moderate correlation with SASSAD. The poor correlation between clinical severity and patients' quality of life is a well-recognised issue, and has been shown in other chronic diseases such as pemphigus.[48,49] It is possible that the two types of scores reflected different facets of AD. The discrepancy highlights the importance of paying equal attention to both elements and

collecting both scores in holistic patient management and clinical trials. Repeating this study with a larger sample size may be required to further investigate this.

An issue raised during study feedback was the responsiveness of the combinative approach used by all of these scoring systems to obtain the score. Severity of the lesions is expressed as a mix of both the reversible disease activity and the more chronic damage, possibly confounding the physician's evaluation of a patient's responsiveness to change on follow-up. When chronic signs such as lichenification have more weighting, the detection of the effect of the current management regime in certain patients may be misperceived. This could possibly be addressed if these scores were modified to assess the patient's activity and secondary damage separately. This approach has already been utilised in other dermatological conditions such as the Cutaneous Lupus Erythematosus Disease Area and Severity Index (CLASI) for lupus and the Pemphigus Disease Area Index (PDAI) for pemphigus.[50,51]

One limitation of this study was its relatively small sample size. Ideally the scoring sessions should be repeated with a greater number of participants of 15-20 to allow for more previse confidence intervals. However, based on the results gathered from the study, it is possible to obtain discriminating results regarding the inter-rater and intra-rater reliability of oSCORAD, EASI, SASSAD and TIS with this sample size. Also, a larger sample size would introduce assessor fatigue given that each assessor had to score each patient using 4 different scores, plus another 2 scores repeated for intra-rater reliability testing. Another limitation was the introduction of potential recall bias by having each assessor completing one score after another which may possibly confound the results of the later score sets.

This study implicates further follow-up studies on the inter-rater and intra-rater reliability of these outcome measures in other countries, settings and skin colours, which led to the second and third Chapters of this thesis. Also, a study to assess the scoring systems for patients over a period of time may be done to assess changes in disease activity and damage. Although more time consuming, this approach would provide valuable information regarding the scoring systems' responsiveness to patients' progress.

Chapter 3. Virtual study of AD outcome measures for skin of colour patients

Aim

To compare the reliability and convergent construct validity of EASI, oSCORAD, TIS and SASSAD in patients with various levels of skin darkness.

Method

Participants and assessors

The full body photographs of 20 patients with AD were obtained from dermatology outpatient clinics from Sydney. Two patients were later excluded as they had more than 2 body parts missing from their full body photographs. Five assessors participated in the scoring process. All assessors were either qualified dermatologists or have been doing full-time dermatology research, and hence had been familiar with atopic dermatitis. Two post-dermatology fellows had trained in the Philippines and were used to darker skin patients. One (Professor Dedee Murrell) had trained in North Carolina where approximately one third of patients were African-Americans. All assessors were required to attend a training lecture on the use of each of the EASI, oSCORAD, TIS and SASSAD which I gave that morning. Also, the assessors were required to attend a debrief session prior to each scoring session to raise queries regarding the administration of these scoring systems. The assessors were completely blinded to the identity of the patients chosen.

Scoring sessions

The assessments were performed over two hour sessions, over four separate days at the weekly departmental research meeting where all assessors viewed the same photographs at the same time. Each session was limited to two hours in length, to avoid assessor fatigue. Full

body photographs of the 18 patients were presented on a screen of at least 1.5 metres by 1.5 metres. Three patients with various levels of skin pigmentations, with unknown identity to the assessors, were also arranged by a separate investigator to have their photographs repeatedly shown at the end for intra-rater reliability testing. For each of the patients scored, the assessors were given four colour-coded scoring sheets including the four measures. The assessors were given the time to view the photographs until they were satisfied with their scores. Each assessor was neither allowed to look at their own scores from the other outcome measures, nor another assessor's scores. They weres seated with at least one empty seat between them. When any patients had minor body parts missing, which five patients did, all assessors were asked not to assess the particular missing body part across all scores.

Categorisation of skin darkness

Each patient's skin pigmentation was scored by all assessors on a numerical scale of 0 to 10, ranging from 0 representing no pigmentation, to 10 representing the darkest level of pigmentation. The average of each patient's pigmentation score across the five assessors was then used to categorise patients into three groups: non-pigmented (score range 0-3), mildly pigmented (score range 3.1-7) and highly pigmented (score range 7.1-10). These ranges were chosen as they approximately divide 0-10 into three equal categories.

Statistical analysis

All outcome measure scores were calculated by two separate study investigators. Data input was performed by one investigator, then, separately double checked by another investigator. The five patients with minor body parts missing from their photographs had their EASI,

SASSAD and oSCORAD's total denominators reduced to reflect the exclusion of the corresponding body parts.

All statistical analyses were performed using SPSS Version 22.0. For reliability testing, both inter-rater and intra-rater reliabilities were assessed by the ICC with 95% CI, using an one-way random analysis variance model, same as the previous study from Chapter 2.

To determine whether the erythema components contributed to the variability in reliability, the erythema component and the 'total minus erythema component' of each outcome measure were separately inputted. The ICCs and CV means of the erythema component and the 'total minus erythema component' were then calculated. The null hypothesis is that the ICC for the erythema component would be bigger than ICC for the 'total minus erythema component', while the CV means across all patients should be smaller. This would indicate that erythema did not contribute to the variability of the scores. To determine the significance of the difference in the ICCs, the CVs of the erythema components and the 'total minus erythema components' were correlated using the paired T-test. This is the same method we used in Chapter 2 for examining the contribution of BSA to the score variability.

For convergent construct validity, all outcome measures were correlated with each other, and the respective Spearman rho correlation coefficients were determined.

Additional assessments by an overseas dermatologist

Additional convergent construct validity testing was performed by a South African dermatologist with expertise in assessing atopic dermatitis in pigmented skin patients

(N.C.D). The results were compared to the Australian dermatologist's results to determine whether experience would improve the validity of outcome measures in highly pigmented skin patients.

Results

Patients demographics

Altogether 18 patients were included in the final analysis for inter-rater reliability and convergent construct validity (Table 4). Out of these, 3 were of Asian background, 11 were of Caucasian background, 2 were of African background and 4 were of Indian background. Three patients were also used for intra-rater reliability analysis (Table 5).

Table 4. Characteristics of the total patient cohort (n=18).

	Highly pigmented*	**Mildly pigmented**	**Non-pigmented**	**Overall**
N (%)	4 (22.2)	7 (38.9)	7 (38.9)	18 (100)
Age range	7-40	2-65	1-65	1-65
Mean age	26.8	29.1	32.3	29.8
Male (%)	2 (50)	5 (71.4)	3 (42.9)	10 (55.6)
Female (%)	2 (50)	2 (28.6)	4 (57.1)	8 (44.4)

*The groups were derived from the mean pigmentation score given by the assessors: non-pigmented (score range 0-3), mildly pigmented (score range 3.1-7) and highly pigmented (score range 7.1-10).

Table 5. Characteristics of the patients for intra-rater reliability testing (n=3).

	Patient 1	**Patient 2**	**Patient 3**
Age*	60	22	25
Sex	Female	Female	Female
Ethnicity	Caucasian	Asian	African
Average pigmentation score	2.0	5.4	7.6

*When the investigator was unable to ascertain what the patient's age was when his/her photo was taken, an estimate of age was made by 2 separate investigators.

Intra-rater reliability

For the patient with highly pigmented skin, the intra-rater reliability ICCs were poor in all scores except for the SASSAD (Table 6). For the patient with mildly pigmented skin, the intra-rater reliability ICCs were poor in the EASI and oSCORAD, and fair for TIS and SASSAD. For the patient with non-pigmented skin, the intra-rater reliabilities of the oSCORAD and TIS were poor, but EASI and SASSAD were good. Given the wide 95% CI of the results, the findings were limited by poor power. However, these results may suggest that in patients with highly pigmented skin, intra-rater unreliability is more likely to be unreliable.

Table 6. Intra-rater reliability ICCs of the EASI, oSCORAD, TIS and SASSAD by pigmentation level.

	EASI	**oSCORAD**	**TIS**	**SASSAD**
Highly pigmented	0.391 (-0.528-0.911)	0.075 (-0.728-0.832)	-0.434 (-0.899-0.574)	0.787 (0.064-0.975)
Mildly pigmented	0.037 (-0.746-0.819)	-0.537 (-0.922-0.476)	0.429 (-0.494-0.918)	0.458 (-0.467-0.924)
Non-pigmented	0.830 (0.184-0.980)	-0.151 (-0.818-0.747)	-0.760 (-0.922-0.122)	0.699 (-0.134-0.963)

Inter-rater reliability

The inter-rater reliability ICCs (and their 95% confidence intervals) for the highly pigmented patients were: TIS -0.21 (-0.24-0.147), SASSAD -0.071 (-0.2-0.631), EASI -0.054 (CI:-0.2-0.657) and oSCORAD -0.089 (-0.206-0.598), indicating very poor inter-rater reliabilities. The ICCs for the mildly pigmented patients were: TIS 0.524 (0.2-0.865), SASSAD 0.341 (0.045-0.775), EASI 0.464 (0.14-0.839) and oSCORAD 0.588 (0.265-0.890), indicating fair inter-rater reliability. The ICCs for the non-pigmented patients were: TIS 0.403 (0.09-0.809), SASSAD 0.667 (0.358-0.916), EASI 0.64 (0.33-0.908) and oSCORAD 0.586 (0.263-0.889),

indicating fair inter-rater reliability for the TIS and oSCORAD, and good inter-rater reliability for the EASI and SASSAD (Table 7).

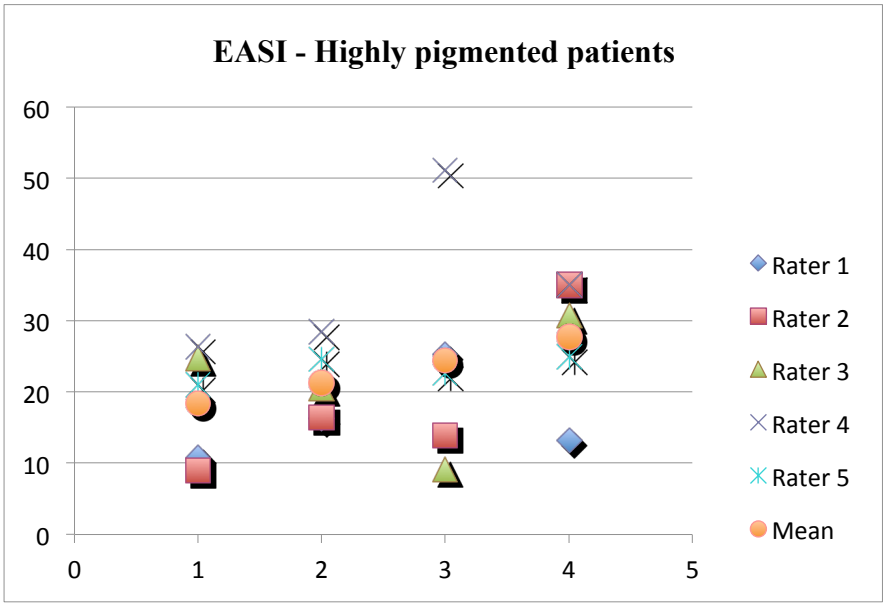
Table 7. Inter-rater reliability ICCs of the EASI, OSCORAD, TIS and SASSAD total scores, the erythema component and the 'total minus erythema component' by pigmentation level.

	EASI	**oSCORAD**	**TIS**	**SASSAD**
Highly pigmented group ICCs (95% CI) (n=4)				
Total score	0.509 (0.300-0.731)	-0.089 (-0.206-0.598)	-0.210 (-0.240-0.147)	-0.071 (-0.201-0.631)
Erythema	-0.171 (-0.200-0.363)	-0.230 (-0.245-0.015)	-0.055 (-0.200-0.656)	-0.088 (-0.206-0.599)
Total minus erythema	-0.072 (-0.201-0.628)	-0.062 (-0.198-0.645)	-0.229 (-0.308-0.399)	-0.144 (-0.222-0.462)
Mildly pigmented group ICCs (95% CI) (n=7)				
Total score	0.464 (0.140-0.839)	0.588 (0.265-0.890)	0.524 (0.199-0.865)	0.341 (0.045-0.775)
Erythema	0.607 (0.290-0.896)	0.636 (0.320-0.906)	0.603 (0.281-0.895)	0.601 (0.280-0.895)
Total minus erythema	0.358 (0.570-0.785)	0.500 (0.176-0.855)	0.413 (0.100-0.814)	0.193 (-0.051-0.670)
Non-pigmented group ICCs (95% CI) (n=7)				
Total score	0.640 (0.330-0.908)	0.586 (0.263-0.889)	0.403 (0.092-0.809)	0.667 (0.358-0.916)
Erythema	0.749 (0.470-0.941)	0.423 (0.108-0.820)	0.409 (0.097-0.812)	0.783 (0.526-0.950)
Total minus erythema	0.533 (0.209-0.869)	0.560 (0.235-0.879)	0.303 (0.180-0.751)	0.459 (0.139-0.837)
Overall ICCs (95% CI) (n=18)				
Total score	0.431 (0.223-0.673)	0.621 (0.423-0.805)	0.480 (0.270-0.710)	0.504 (0.294-0.747)
Erythema	0.609 (0.409-0.798)	0.509 (0.300-0.731)	0.568 (0.363-0.772)	0.701 (0.522-0.853)
Total minus erythema	0.509 (0.300-0.731)	0.597 (0.396-0.790)	0.436 (0.228-0.677)	0.340 (0.141-0.598)

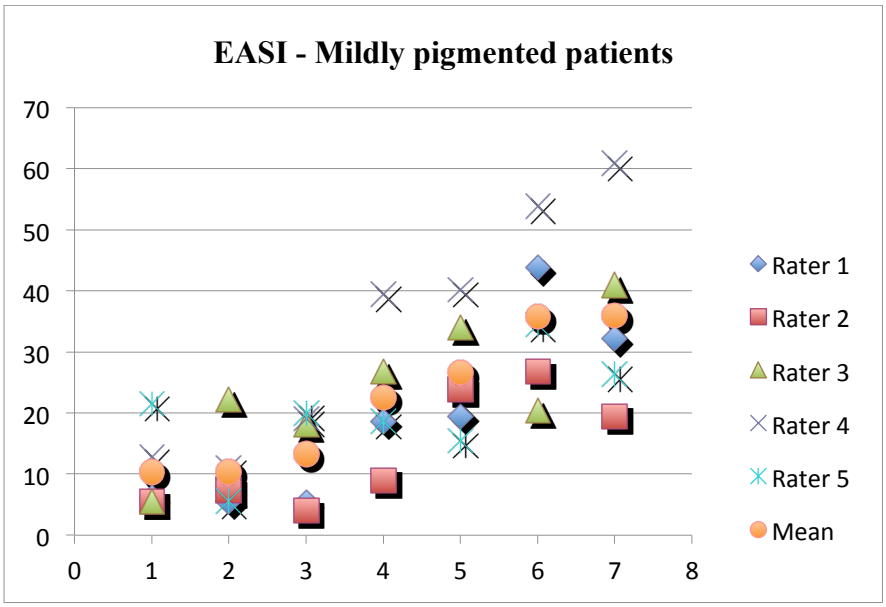
Inter-rater scatterplots showed that in highly pigmented skin patients, all scores have poor inter-rater reliability regardless of disease severity. In mildly pigmented patients, all scores had poorer inter-rater reliability with increased disease severity. In non-pigmented patients, inter-rater reliability does not appear influenced by disease severity (Figure 19).

Figure 19. Inter-rater scatterplots demonstrating the spread of scores for all 4 outcome measures in patients of various pigmentation levels. EASI scores for a) Highly pigmented patients, b) Mildly pigmented patients and c) Non-pigmented patients; oSCORAD scores for d) Highly pigmented patients, e) Mildly pigmented patients and f) Non-pigmented patients; SASSAD scores for g) Highly pigmented patients, h) Mildly pigmented patients and i) Non-pigmented patients; TIS scores for j) Highly pigmented patients, k) Mildly pigmented patients and l) Non-pigmented patients. These scores demonstrate that in patients with highly pigmented skin, all scores and across all disease severity have poor inter-rater reliability. In patients with mildly pigmented skin, across all scores, inter-rater reliability is decreased as disease severity is increased. In patients with non-pigmented skin, generally inter-rater reliability is not influenced by disease severity.

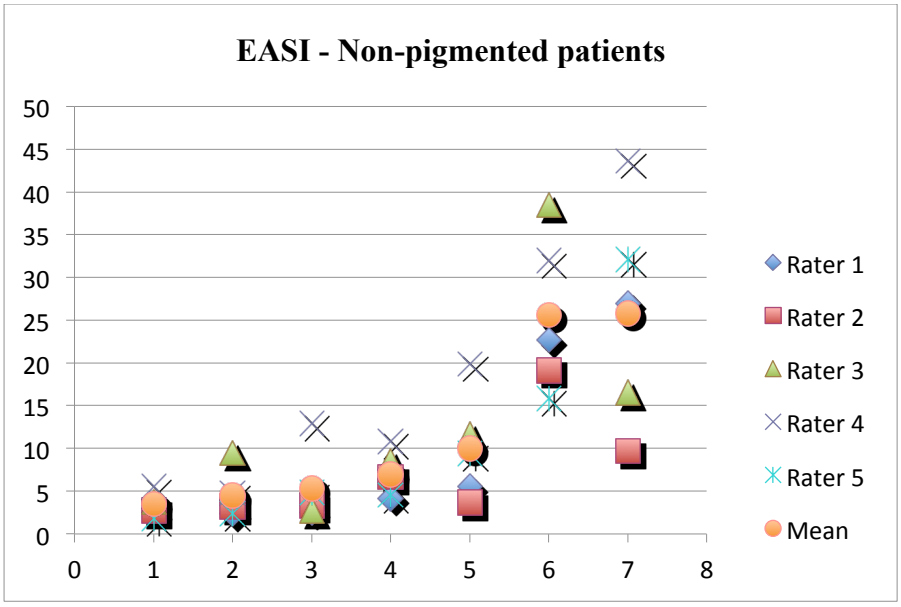
a)



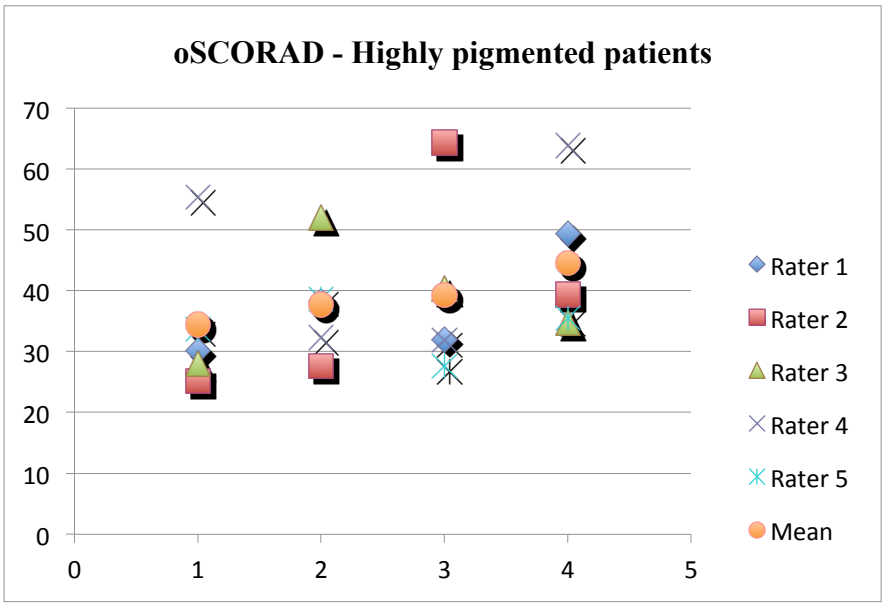
b)



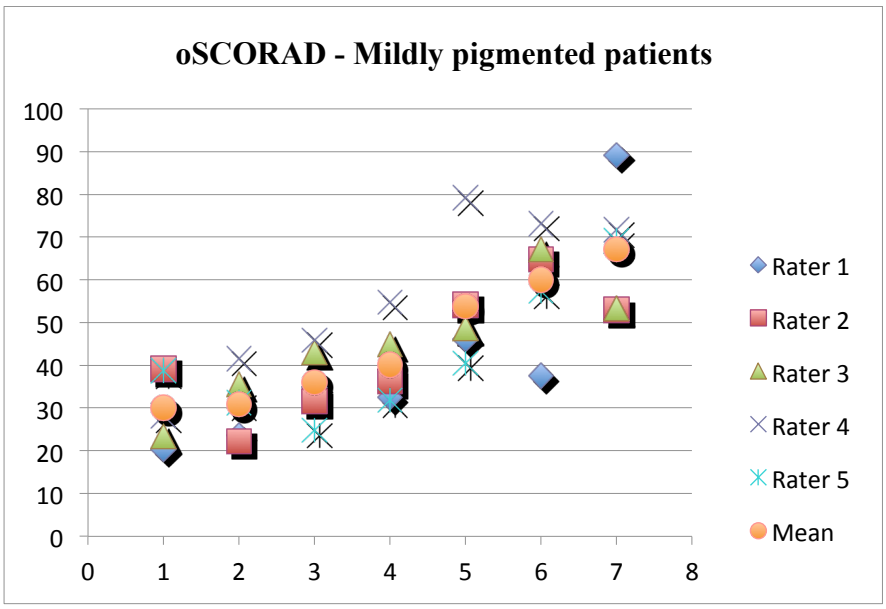
c)


EASI - Non-pigmented patients
50
45
40
35
30
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8
Rater 1
Rater 2
Rater 3
Rater 4
Rater 5
Mean

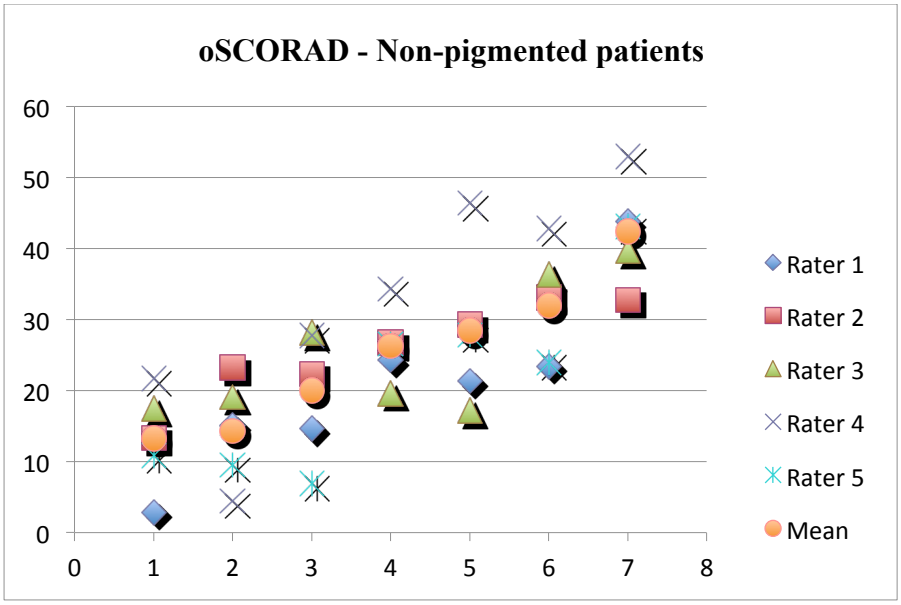
d)


oSCORAD - Highly pigmented patients
70
60
50
40
30
20
10
0
0 1 2 3 4 5
Rater 1
Rater 2
Rater 3
Rater 4
Rater 5
Mean

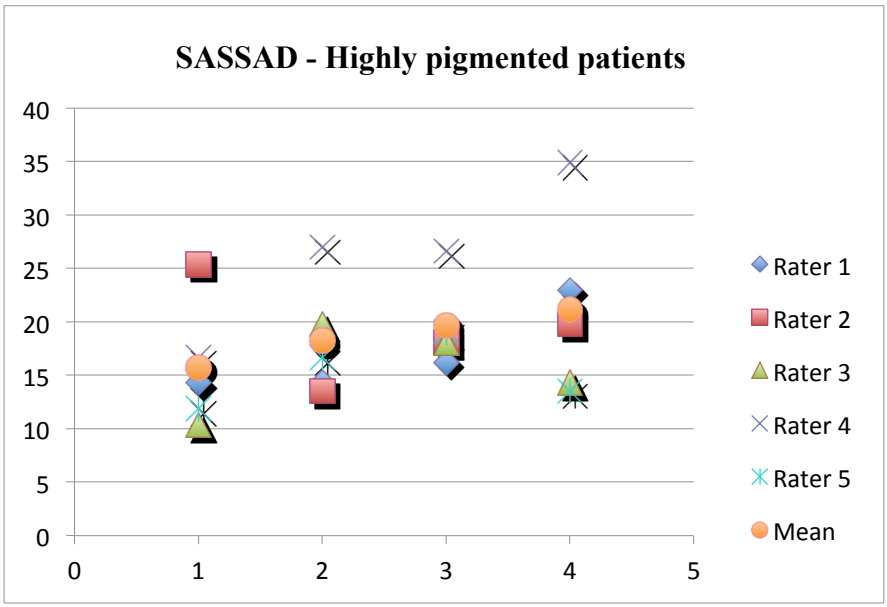
e)


oSCORAD - Mildly pigmented patients
100
90
80
70
60
50
40
30
20
10
0
0 1 2 3 4 5 6 7 8
Rater 1
Rater 2
Rater 3
Rater 4
Rater 5
Mean

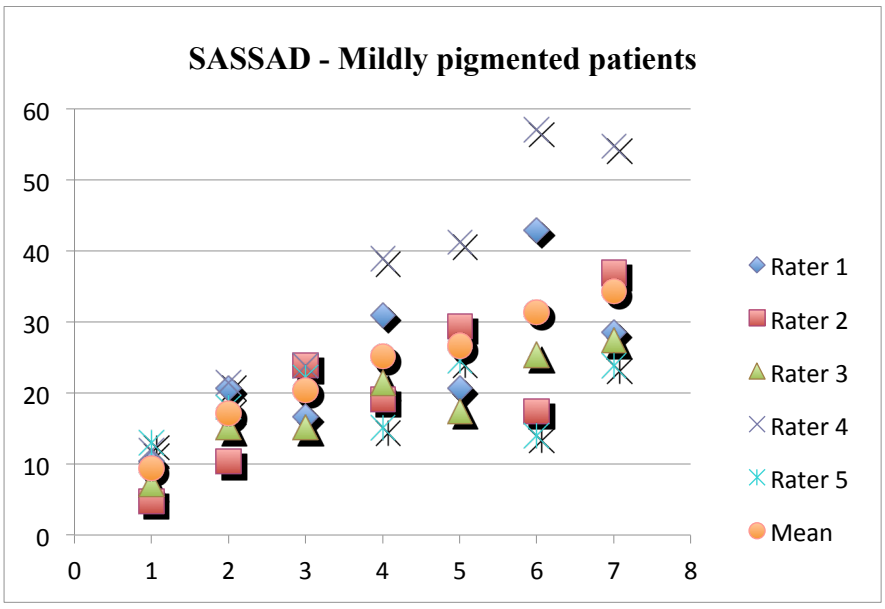
f)


oSCORAD - Non-pigmented patients
60
50
40
30
20
10
0
0 1 2 3 4 5 6 7 8
Rater 1
Rater 2
Rater 3
Rater 4
Rater 5
Mean

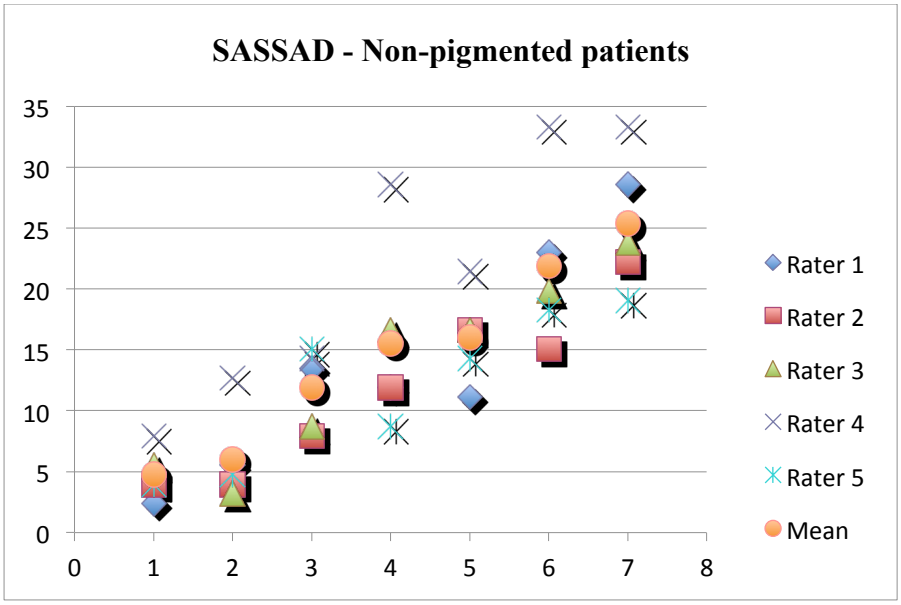
g)


SASSAD - Highly pigmented patients
40
35
30
25
20
15
10
5
0
0 1 2 3 4 5
Rater 1
Rater 2
Rater 3
Rater 4
Rater 5
Mean

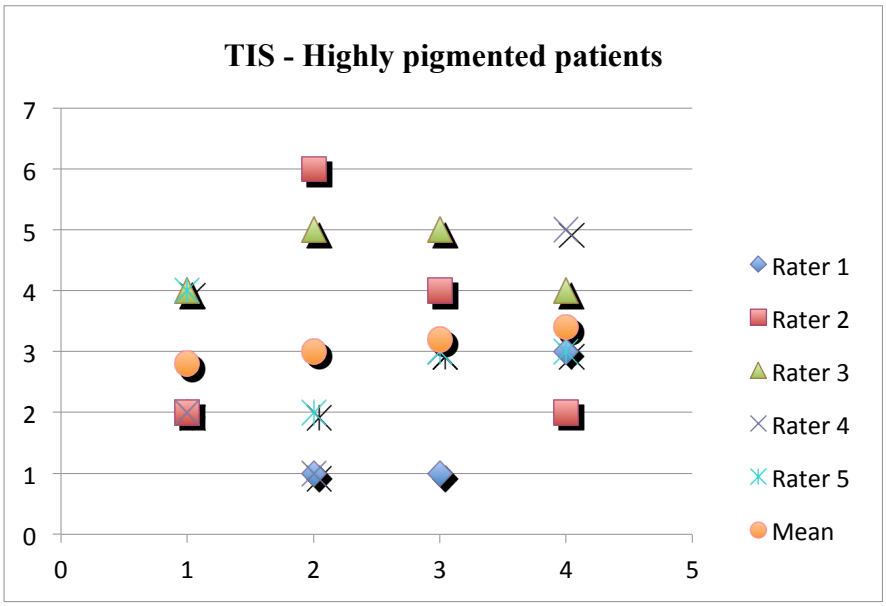
h)


SASSAD - Mildly pigmented patients
60
50
40
30
20
10
0
0 1 2 3 4 5 6 7 8
Rater 1
Rater 2
Rater 3
Rater 4
Rater 5
Mean

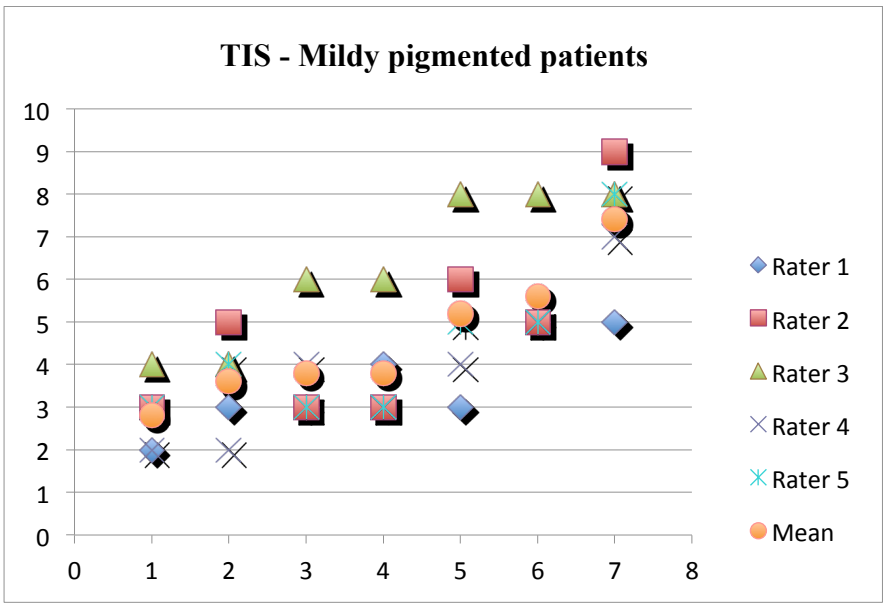
i)


SASSAD - Non-pigmented patients
35
30
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8
Rater 1
Rater 2
Rater 3
Rater 4
Rater 5
Mean

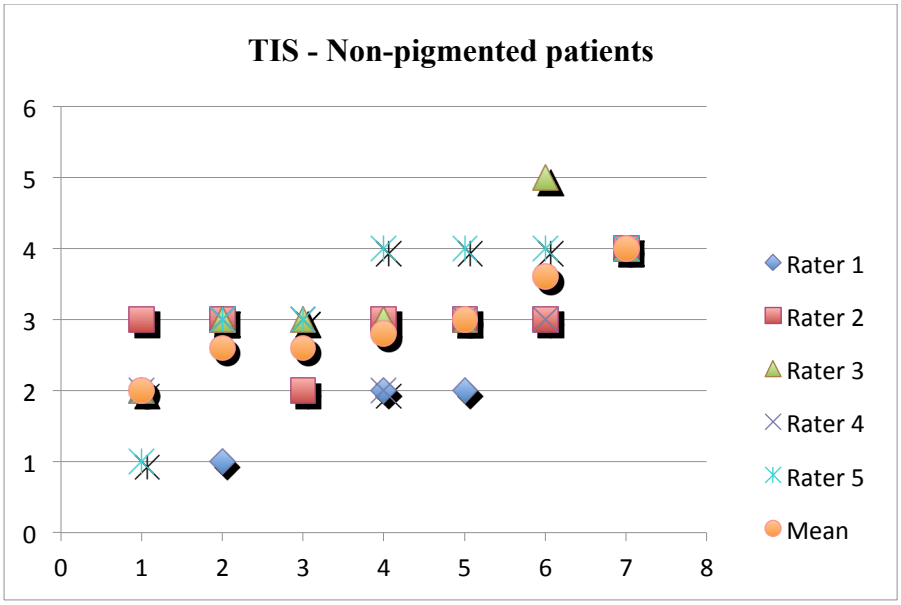
j)


TIS - Highly pigmented patients
7
6
5
4
3
2
1
0
0 1 2 3 4 5
Rater 1
Rater 2
Rater 3
Rater 4
Rater 5
Mean

k)


TIS - Mildy pigmented patients
10
9
8
7
6
5
4
3
2
1
0
0 1 2 3 4 5 6 7 8
Rater 1
Rater 2
Rater 3
Rater 4
Rater 5
Mean

l)



Erythema's contribution to variability

The inter-rater reliability ICCs and coefficient of variations (CV) means of the erythema components and 'total minus erythema components' were compared against each another. For the highly pigmented patients, the EASI and oSCORAD had slight, but clinically significant superior reliabilities when the erythema components were excluded (EASI -0.171 vs. -0.072, p=0.034; oSCORAD -0.230 vs. -0.062, p=0.04) (Table 7). This superior reliability was not present in the TIS or SASSAD, or in any other pigmentation groups. These results suggest that the erythema components have likely contributed to the variability of EASI and oSCORAD in highly pigmented skin patients. However, given the poor ICCs of the 'total minus erythema components', other factors were likely to have also contributed to the variability.

When comparing the CV means (an indicator of variability) in highly pigmented patients, there were higher CV means in the erythema components than the 'total minus erythema component', across all outcome measures (Table 8). The higher CV mean of the erythema

component was not present in any other outcome measure or other pigmentation groups, except for SASSAD in mildly pigmented patients. These results were also evident of erythema's contribution to the variability of the outcome measures in highly pigmented patients.

Table 8. Coefficient of variation (CV) means of the erythema components versus the 'total minus erythema components' of the 4 outcome measures by pigmentation level.

	EASI	**oSCORAD**	**TIS**	**SASSAD**
Highly pigmented group CVs (n=4)				
Erythema	0.987	0.961	1.01	0.852
Total minus erythema	0.525	0.331	0.499	0.427
Mildly pigmented group CVs (n=7)				
Erythema	0.359	0.323	0.249	0.323
Total minus erythema	0.659	0.283	0.377	0.471
Non-pigmented group CVs (n=7)				
Erythema	0.578	0.284	0.201	0.387
Total minus erythema	0.746	0.526	0.807	0.647

Convergent construct validity

In highly pigmented patients, none of the scoring instruments significantly correlated with each other. In mildly pigmented patients, SASSAD is not statistically significantly correlated with any of the other scoring instruments. The correlations of the other 3 outcome instruments were EASI with TIS: 0.829 (p=0.021), EASI with oSCORAD: 0.857 (p=0.014) and oSCORAD with TIS: 0.919 (p=0.003). In non-pigmented patients, the correlations of all 4 outcome measures were mostly statistically significant: EASI with TIS: 0.919 (p=0.003), EASI with oSCORAD: 1.000 (p<0.01), EASI with SASSAD: 0.786 (p=0.036), oSCORAD with TIS: 0.919 (p=0.003), oSCORAD with SASSAD: 0.786 (p=0.036) and TIS with SASSAD: 0.793 (p=0.033). These results suggest the poorer convergent construct validity of all outcome measures in highly pigmented patients.

Additional assessments by an overseas dermatologist

The same 18 patients were assessed by the South African dermatologist. In highly pigmented patients, none of the scoring instrument significantly correlated, except for EASI with SASSAD (p=0.051). In mildly pigmented patients, all outcome measures correlated well, but was limited by low power with only 2 patients in this group (as per pigmentation score out of 10 rated by the South African dermatologist). In non-pigmented patients, the correlations of all 4 outcome measures were mostly statistically significant: EASI with TIS: 0.645 (p=0.023), EASI with oSCORAD: 0.867 (p<0.01), EASI with SASSAD: 0.967 (p<0.01), oSCORAD with TIS: 1.000 (p<0.01), oSCORAD with SASSAD: 0.872 (p<0.01) and TIS with SASSAD: 0.648 (p=0.023). These results were comparable to the Australian dermatologist, confirming the poorer convergent construct validity of all outcome measures in highly pigmented patients.

Discussion

This is the first study to evaluate the outcome measures for AD clinical signs in patients with dark skin. Overall, all AD outcome measures, the EASI, the oSCORAD, the TIS and the SASSAD may have poor reliability and validity in patients with very dark skin. All measures had very poor inter-rater reliability when used in patients of highly pigmented skin with ICCs of <0.4, regardless of disease severity.

The study suggested that erythema may be a contributor to the inter-rater variations with higher CV means of erythema components compared to the 'total minus erythema components'. This finding echoed a previous study by Ben-Gashir and Hay, which investigated the effect of including and excluding erythema assessment on the assessment of

AD severity; when the erythema component of the oSCORAD was excluded and adjusted for the total, disease was found to be severer in black children.[52] Our study went into more breadth to evaluate this interesting question. Given that in our study, the inter-rater ICCs of the 'total minus erythema components' were also poor in the highly pigmented patients, other factors are likely have also contributed to the variability.

Several key weaknesses limited the validity of this study. Firstly, the inferior reliability of all outcome measures in lighter patients suggested by this study when compared to other studies is likely due to the fact that patients were virtually scored instead of in a routine clinic setting. Three-dimensional assessments for lichenification and papulation were likely compromised, as was the feeling of heat via palpation to distinguish between active inflammation and post-inflammatory hyperpigmentation. In addition, photos were taken with different cameras and by different photographers, leading to variations in photo quality. We also acknowledge the limited number of patients used for this study, given the difficulties to obtain high definition, full body photos of very dark patients. Another limitation was that being a single-centre study, the poorer reliability and validity may not be demonstrated in all other setting and locations. Lastly, this study could be improved if the categorisation of skin darkness was performed more objectively, for example, using a mexameter or spectrophotometer.

All four measures had inferior construct validity in highly pigmented patients, even when scored by a South African dermatologist with experience in assessing dark skin. This is suggestive of underlying heterogenicity of clinical signs in highly pigmented patients, which may not have been included in the four common AD outcome measures.

From this study, a new outcome measure may be needed to take into account the manifestation and perception of erythema, as well as other clinical heterogeneity found in dark AD patients such as ichthyosis or prurigo nodularis.[53] A grey-scale in addition to or instead of the EASI's erythema scale may more accurately reflect the erythema shown in ethnic skin, while a heat or temperature scale may help to distinguish between inflammation and post-inflammatory hyperpigmentation. This study warranted another improved comparison study in real-life with a larger sample size and an objective method of categorising patient skin colours.

Chapter 4. Multi-centre study of AD outcome measures for skin of colour patients

Aims

- To improve upon the previous study (Chapter 3) and compare the inter-rater and intra-reliability of the EASI and oSCORAD.
- To investigate whether erythema was contributing to the inter-rater variability in scoring.
- To investigate whether a novel grey-scale instead of erythema scale would be more reliable for dark skinned patients than the erythema component of the EASI.
- To investigate if there are any significant correlations between patient-reported symptoms and clinician-rated signs in patients with dark skin.

Method

This prospective study was granted ethical approval from the Bellberry Human Research Ethics Committee, Australia (protocol ID: 2015-06-445). The scoring was conducted on 25th October 2015 in Sydney and 29th of November 2015 in Melbourne by the same team of doctors.

Sample size calculations

The sample size was calculated in the same way as per the study in Chapter 2. Each skin colour group required 11 patients and the same 5 assessors.

Participants and assessors

Patients from outpatient clinics were recruited. An excel sheet was used to stratify the recruited patients and ensure that were at least 12 patients with the skin types I to III and

another 12 patients with skin types IV to VI. Patients of all ages and AD severity were recruited. The inclusion and exclusion criteria is outlined in Table 9.

Table 9. Inclusion and exclusion criteria for the study.

Inclusion criteria	Exclusion criteria
• Age range: 0-80 years • Those with AD, as per Hanifin and Rajka's criteria • Of mature mind and able to give consent, in case of children (below 18 years of age) then the guardian or caretaker can provide consent • Be able to understand and complete the DLQI/ CDLQI and POEM questionnaire (except for babies and infants)	• Participants who are unwilling or unable to comply with requirements of the protocol • Those withdrawing from the study • Non-English speakers

Five assessors participated in all patient assessments in Sydney and Melbourne. Four of the five assessors were experienced dermatologists who regularly see AD patients (Professor Dedee F. Murrell, Professor John Su, Dr Michelle Rodrigues and Dr Linda Martin). One of the assessor was a third year dermatology registrar, Dr Benjamin Daniel (equivalent of resident in the U.S.) who regularly sees AD patients and had previously completed a two-year dermatology fellowship. All assessors were required to attend a 30-minute training lecture on the administration of the outcome measures, in particular the new grey-scale, and raise any queries prior to each of the scoring session. The assessors were completely blinded to the identity of the patients chosen and the ordering of outcome measures investigated. The skin types of the assessors were: II (D.F.M), III (J.S), IV (M.R), IV (B.D) and I (L.M).

Two independent half-day scoring sessions

Two half-day scoring sessions were conducted, one based in Sydney and one in Melbourne. On the scoring day, participants had the melanin index of their outer gluteal skin measured

using a Mexameter® MX-18 (Courage - Khazaka Electronic, Köln, Germany) in a designated private room by a non-scoring investigstor. Then, each participant was placed in individual rooms and asked to complete patient-oriented questionnaires. To assess inter-rater reliability, each assessor followed a roster to rotate around the rooms. The assessors had 8 minutes each to examine each patient using the 4 scales. Each scale was printed on different coloured papers and placed in a random order for each assessor, per patient. At any one point, only one assessor was assigned to one patient. The assessors were not allowed to discuss their scores with one another. To assess intra-rater reliability, each assessor rescored 10 randomly allocated patients (5 patients with lighter and 5 with darker skin) The randomisation process was kept undisclosed by an uninvolved investigator. There was at least a half an hour time lapse between the first and second assessment to minimise recall bias.

Categorisation of skin darkness

Each participant's skin pigmentation level was categorised based on their objective melanin index on a numerical scale of 0 to 999. Two groups were formed with lighter skin having melanin indices of 0-199 and darker skin having melanin indices of 200-999. The typical phototype and their melanin content as per the mexameter manufacturer guide are outlined in Table 10.

Table 10: Mexameter phototype and melanin content.

Study group	Phototype	Description	Average melanin content measured by melanin index
Lighter skin	I	Celtic	0-49
	II	Caucasian	50-99
	III	European mixed	100-149
	IV	Mediterranean	150-199
Darker skin	V	Asian/Indian	200-299
	VI	Black	>300

Statistical analysis

Data input into a pre-designed excel document was performed by one investigator (myself), then, separately double-checked by another investigator (E.Y and D.D.O). All statistical analyses were performed using SPSS Version 22.0. For reliability testing, both inter-rater and intra-rater reliabilities were assessed by the ICC, using a one-way random analysis variance model. To determine whether erythema scoring contributed to the variability in inter-rater reliability, the 'erythema component' and the 'total minus erythema component' and their respective CV of the EASI and oSCORAD were separately calculated. The CVs of the erythema components and the 'total minus erythema components' were compared using the paired T-test, using the same method as the virtual study in Chapter 3. Correlations between the POEM score and objective outcome measures were analysed using the Spearman's rho with a significance level of 5%.

Results

Patient demographics

In total, 25 participants were included in the final analysis for inter-rater reliability and 6 of these were included for intra-rater reliability testing. Of these, 11 were categorised as having "lighter skin" and 14 were categorised as having "darker skin". Table 11 shows the participants' demographic characteristics. Table 12 shows the mean value of each outcome measure score in each skin colour group.

Table 11. Demographic details of the 25 participants in the study.

Baseline characteristics		Participants
Basic demographic information	Total	25
	Paediatric/adult	13 (52%) /12 (48%)
	Female /male	7 (28%) /18 (72%)
Ethnicity	Caucasian	9 (36%)
	Asian	12 (48%)
	Indian	3 (12%)
	African	1 (4%)
Melanin index	<150	4 (16%)
	150-199	7 (28%)
	200-400	11 (44%)
	>400	3 (12%)
Patient-oriented severity scores according to POEM bands*	Almost clear	1 (4%)
	Mild	5 (20%)
	Moderate	14 (56%)
	Severe	3 (12%)
	Very severe	2 (8%)

*POEM score severity bands: 0–2 (clear/almost clear); 3–7 (mild); 8–16 (moderate); 17–24 (severe); 25–28 (very severe)

Table 12. Mean scores (if clinician-rated, across all scorers) and range of each outcome measure by skin colour group.

	EASI	oSCORAD	IGA	POEM	DLQI
Lighter skin	10.9 (1.4-45.2)	26.5 (7.1-50.1)	2.6 (1.2-4.8)	15.4 (4-19)	9.8 (2-28)
Darker skin	12.4 (0.6-41.7)	26.7 (10.4-53.0)	2.5 (1.8-4.2)	9.4 (2-28)	10.7 (1-19)

Intra-rater reliability

All scores demonstrated excellent intra-rater reliability in both lighter skin and darker skinned participants with very high levels of ICCs, with the exact data shown in Table 13.

Table 13. Intra-rater reliability scores by skin colour group.

	EASI	oSCORAD	IGA	Greyscale
Lighter	0.944 (0.879-0.975)	0.872 (0.734-0.941)	0.902 (0.792-0.955)	
Darker	0.957 (0.906-0.981)	0.958 (0.909-0.981)	0.888 (0.765-0.949)	0.912 (0.813-0.960)

Inter-rater reliability

For inter-rater reliability in "lighter skin" participants, both EASI (ICC=0.827 [95%CI: 0.658-0.941]), and IGA (ICC=0.803 [95%CI: 0.618-0.932]) demonstrated excellent correlations, whilst oSCORAD (ICC=0.680 [95%CI: 0.441-0.880]) demonstrated good correlations. For inter-rater reliability in "darker skin" participants, only the EASI (ICC=0.774 [95%CI: 0.598-0.906]) demonstrated excellent correlation, whilst the oSCORAD (ICC=0.736 [95%CI: 0.544-0.889]), IGA (ICC=0.696 [95%CI: 0.490-0.868]) and grey-scale (ICC=0.638 [95%CI: 0.400-0.838]) demonstrated good correlations. We also calculated the ICC of the EASI with the grey-scale replacing its erythema scale, which showed an excellent ICC of 0.776 (95%CI: 0.601-0.907).

To explore the confounding effects of age and severity on the objective outcome measures' inter-rater reliabilities, we also calculated the inter-rater ICCs of EASI, oSCORAD and IGA (adult versus children) and by POEM severity scores (the lower 50 percentiles versus the higher 50 percentiles). The results of these are outlined in Table 14.

Table 14. The inter-rater ICCs (and 95% CI) of the EASI, oSCORAD and IGA scores in AD participants by age and patient-oriented POEM score.

	EASI	**oSCORAD**	**IGA**
Adult	0.789 (0.605-0.922)	0.743 (0.537-0.902)	0.772 (0.579-0.914)
Paediatric	0.710 (0.500-0.811)	0.610 (0.375-0.829)	0.709 (0.499-0.880)
Milder 50 percentile of patients by POEM severity (N=13)	0.644 (0.416-0.848)	0.541 0.298-0.789)	0.317 (0.137-0.674)
More severe 50 percentile of patients by POEM severity (N=12)	0.788 (0.603-0.921)	0.703 (0.481-0.883)	0.814 (0.645-0.932)

Erythema's contribution to variability

The CV of the erythema components and the 'total minus erythema components' were compared against each other for the EASI and oSCORAD (Table 15). For the light skin group, the EASI score had a significantly higher CV when the erythema component was omitted compared to erythema component alone (mean CV value: 0.57 versus 0.37, Paired T-test, p=0.001). This means that when EASI score's erythema component is removed in lighter skin patients, the score became more variable between assessors. This did not apply for the EASI score with the darker skinned patients or the oSCORAD. The EASI minus erythema for the dark skinned patients gave a CV value of 0.47 and while the grey scale had a CV value of 0.66.

Table 15. Coefficient of variation (CV) means of the erythema components versus the 'total minus erythema components' of the EASI and SCORAD.

	EASI - CV	**EASI - ICC**	**oSCORAD - CV**	**oSCORAD - ICC**
Lighter skin				
Erythema	0.367	0.837 (0.674-0.945)	0.243	0.482 (0.218-0.776
Total minus erythema	0.571	0.796 (0.607-0.929)	0.316	0.691 (0.454-0.885)
Paired T-test	p=0.001*	-	p=0.335	-
Darker skin				
Erythema	0.400	0.800 (0.637-0.918)	0.269	0.416 (0.184-0.696)
Total minus erythema	0.416	0.746 (0.558-0.893)	0.295	0.719 (0.520-0.880)
Paired T-test	p=0.467	-	p=0.233	-

Correlation between POEM and clinician-reported measures

In darker skinned patients, the POEM showed no correlation with the EASI (rho 0.53, p=0.05) or oSCORAD (rho 0.42, p=0.14), but showed a correlation with the IGA (rho 0.64, p=0.013). In lighter skinned patients, the POEM showed no correlation with the EASI (rho 0.48, p=0.14), oSCORAD (rho 0.31, p=0.35) or IGA (rho 0.30, p=0.38).

Discussion

Once again, the EASI showed excellent inter-rater and intra-rater reliability in patients of all skin colours. Therefore, we strongly support the HOME consensus that the EASI should be used as the core outcome measure in AD clinical trials all around the world. On the other hand, the oSCORAD showed only good inter-reliability in both skin colour groups, inferior to that of the EASI, question its need as an optional outcome measure in clinical trials.

As opposed to findings from Chapter 3,[39] in darker skinned patients, erythema perception was not found to be a contributor to the inter-rater variability in the EASI or oSCORAD. Surprisingly, in lighter skinned patients, EASI score's erythema component reduced its inter-rater variability. One explanation may be that in lighter skin patients, erythema can be more reliably assessed amongst clinicians than the other signs such as lichenification, papulation/induration and excoriations.

The greyscale for dark skinned patients showed excellent intra-rater reliability and inter-rater reliability when replacing the erythema scale of the EASI, very similar in value to the original EASI. We suspect the reliability of the greyscale may be under-estimated by the study, as the assessors only had a single training session prior to using the greyscale, and were unfamiliar with its usage, unlike the EASI or oSCORAD which they have had more previous clinical experiences with. In very dark skinned patients with minimal erythema detectable, the greyscale may be more useful to substitute for erythema.

Interestingly, the POEM appeared lower in darker skin patients than lighter skin patients while all other scores were similar. This may suggest that patients with darker skin may perceive their symptoms of AD more lightly compared to their lighter skin counterparts,

while their objective signs remain comparable. Statistical analysis showed an overall poor correlation between the patient's clinical signs and symptoms. This highlighted that the two types of scores reflects different facets of AD and the importance of collecting both scores in holistic patient management and clinical trials. Further research may be warranted regarding the exact symptoms affected in dark skinned patients versus lighter skinned with AD patients.

The EASI and oSCORAD's reliabilities were shown to be much higher in our study than our previous study performed using photographs, where the reliability of all clinician-rated AD outcome measures were poor in dark skinned patients, and were either fair or good in light skinned patients.[39] Given the recent advances in teledermatology worldwide, this highlights that when assessing AD patients, it is optimal to perform three-dimensional real life assessments.

One limitation of this study was its relatively small sample size. Ideally, the very dark Fitzpatrick skin type VI patients should be categorised separately from the relatively darker Fitzpatrick skin type IV and V patients, with participants of 15 or more per group to allow for more precise confidence intervals. However, based on the results gathered from the study, it is possible to validate the EASI score as having an acceptable level of reliability for performing clinical trials in dark skinned patients. Also, there may be potential recall bias by having each assessor completing the score after one another which may possibly confound the results of the later score sets, especially with the greyscale and the EASI in the same session. As the method use in the first study in Chapter 2, the authors have tried to minimise this recall bias by putting the score sheets in random order.

Chapter 5. Conclusions

This Master of Science thesis strongly supports the HOME consensus statement recommending EASI as the core measure for the clinical signs of AD in clinical trials all around the world. The first study (Chapter 1) showed that the EASI has highest inter-rater and intra-rater reliability in comparison with the oSCORAD, SASSAD and TIS scores, and would be a valuable and reliable tool for clinical trials. TIS was the easiest to administer and would be useful as a personal marker of patients' progress in general practice, scored by the same doctor each time, but not in situations where there are different doctors each time or in clinical trials. The well-accustomed oSCORAD showed relative poorer inter-rater reliability compared to the SASSAD and EASI, questioning its need as an optional outcome measure in clinical trials. Interestingly, the study also revealed low correlation between patients' perception of illness through quality of life questionnaires and dermatologists' objective assessment of disease severity.

The second study (Chapter 3) prompted the third study (Chapter 4), and highlighted the fact that in order to reliably and accurately assess the clinical severity of AD patients, real-life patients rather than virtual assessment from photographs is mandatory. The third study showed that the EASI has excellent reliability in AD patients of all skin colours, and advocates for the inclusion of dark skinned patients in future AD clinical trials. The greyscale was shown to have similar reliability to the EASI score for dark skinned patients. Patient-oriented symptoms measures did not correlate to clinical-rated outcome measures, and patients with darker skin may perceive their symptoms as less severe to their white counterparts.

Publications / related editorials /conference presentations associated with this thesis

Publications:

- **Zhao CY**, Hao EY, Oh DD, Daniel BS, Martin LK, Su JC, Rodrigues M, Murrell DF. A comparison study of clinician-rated atopic dermatitis outcome measures for intermediate to dark skin patients. British Journal of Dermatology. (Original research) In press 2016 Nov, ePub available online
- **Zhao CY,** Tran AQ, Lazo-Dizon JP, et al. A pilot comparison study of four clinician-rated atopic dermatitis severity scales. The British journal of dermatology 2015;173:488-97.
- **Zhao CY**, Wijayanti A, Doria MC, et al. The reliability and validity of outcome measures for atopic dermatitis in patients with pigmented skin: A grey area. International Journal of Women's Dermatology;1:150-4.

Related editorial:

- Thomas, K. S. (2015). "EASI does it: a comparison of four eczema severity scales." Br J Dermatol 173(2): 316-317.

Oral presentations (presenter underlined):

- **Zhao CY,** Zhao CY, Hao EY, Oh DD, Daniel BS, Martin LK, Su JC, Rodrigues M, Murrell DF. Atopic dermatitis scores for dark skinned patients. *9th Georg Rajka International Symposium on Atopic Dermatitis*, Brazil. May 2016
- **Zhao CY,** Tran AQT, Wijayanti A, Doria MJ, Harris AG, Jain SV, Legaspi K, Dlova NC, Lazo-Dizon JP, Kim J, Daniel BS, Venugopal SS, Rhodes LM, Law MG, Murrell DF,

Atopic dermatitis severity scales. *Australasian Society for Dermatologic Research Meeting*, Adelaide, May 2015.

- **Zhao CY,** Wijayanti A, Doria MC, Harris AG, Jain SV, Legaspi K, Law MG, Murrell, DF. Assessing Atopic Dermatitis in Pigmented Skin Patients (A Grey Area). *Skin of Color Society Annual Meeting,* San Francisco, USA, Mar 2015.

Poster presentations:

- Zhao CY, Hao EY, Oh DD, Daniel BS, Martin LK, Su JC, Rodrigues M, Murrell DF. Atopic dermatitis scores for dark skinned patients. Australasian College of Dermatologists Annual Scientific Meeting, Perth, May 2016
- Zhao CY, Tran AQT, Lazo-Dizon JP, Kim J, Daniel BS, Venugopal SS, Rhodes LM, Law MG, Murrell DF. Comparison Study of Four Outcome Measures for Atopic Dermatitis. At World Congress of Dermatology, Vancouver, Canada, Jun 2015
- Zhao CY, Tran AQT, Lazo-Dizon JP, Kim J, Daniel BS, Venugopal SS, Rhodes LM, Law MG, Murrell DF. A Prospective Comparison Study of Four Outcome Measures for Atopic Dermatitis. Australasian College of Dermatologists Annual Scientific Meeting, Adelaide, May 2015
- Zhao CY, Wijayanti A, Doria MC, Harris AG, Jain SV, Legaspi K, Law MG, Murrell, DF. Assessing Atopic Dermatitis in Pigmented Skin Patients (A Grey Area). Cutaneous Biology Meeting, North Stradbroke Island, September 2014

References

1. Shaw TE, Currie GP, Koudelka CW, Simpson EL. Eczema prevalence in the United States: data from the 2003 National Survey of Children's Health. The Journal of investigative dermatology 2011;131:67-73.

2. Sugiura H, Umemoto N, Deguchi H, et al. Prevalence of childhood and adolescent atopic dermatitis in a Japanese population: comparison with the disease frequency examined 20 years ago. Acta dermato-venereologica 1998;78:293-4.

3. Oh JW, Pyun BY, Choung JT, et al. Epidemiological change of atopic dermatitis and food allergy in school-aged children in Korea between 1995 and 2000. Journal of Korean medical science 2004;19:716-23.

4. Robertson CF, Roberts MF, Kappers JH. Asthma prevalence in Melbourne schoolchildren: have we reached the peak? The Medical journal of Australia 2004;180:273-6.

5. Simpson CR, Newton J, Hippisley-Cox J, Sheikh A. Trends in the epidemiology and prescribing of medication for eczema in England. Journal of the Royal Society of Medicine 2009;102:108-17.

6. Esamai F, Ayaya S, Nyandiko W. Prevalence of asthma, allergic rhinitis and dermatitis in primary school children in Uasin Gishu district, Kenya. East African medical journal 2002;79:514-8.

7. Zar HJ, Ehrlich RI, Workman L, Weinberg EG. The changing prevalence of asthma, allergic rhinitis and atopic eczema in African adolescents from 1995 to 2002. Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology 2007;18:560-5.

8. Deckers IA, McLean S, Linssen S, Mommers M, van Schayck CP, Sheikh A. Investigating international time trends in the incidence and prevalence of atopic eczema 1990-2010: a systematic review of epidemiological studies. PloS one 2012;7:e39803.

9. Schmitt J, Spuls P, Boers M, et al. Towards global consensus on outcome measures for atopic eczema research: results of the HOME II meeting. Allergy 2012;67:1111-7.

10. Schmitt J, Williams H. Harmonising Outcome Measures for Eczema (HOME). Report from the First International Consensus Meeting (HOME 1), 24 July 2010, Munich, Germany. The British journal of dermatology 2010;163:1166-8.

11. Chalmers JR, Schmitt J, Apfelbacher C, et al. Report from the Third International Consensus Meeting to Harmonise Core Outcome Measures for Atopic Eczema / Dermatitis Clinical Trials (HOME). The British journal of dermatology 2014.

12. Vourc'h-Jourdain M, Barbarot S, Taieb A, et al. Patient-oriented SCORAD: a self-assessment score in atopic dermatitis. A preliminary feasibility study. Dermatology (Basel, Switzerland) 2009;218:246-51.

13. Verwimp JJ, Bindels JG, Barents M, Heymans HS. Symptomatology and growth in infants with cow's milk protein intolerance using two different whey-protein hydrolysate based formulas in a Primary Health Care setting. European journal of clinical nutrition 1995;49 Suppl 1:S39-48.

14. Sugarman JL, Fluhr JW, Fowler AJ, Bruckner T, Diepgen TL, Williams ML. The objective severity assessment of atopic dermatitis score: an objective measure using permeability barrier function and stratum corneum hydration with computer-assisted estimates for extent of disease. Archives of dermatology 2003;139:1417-22.

15. Silny W, Czarnecka-Operacz M, Silny P. The new scoring system for evaluation of skin inflammation extent and severity in patients with atopic dermatitis. Acta dermatovenerologica Croatica : ADC 2005;13:219-24.

16. Murao Y, Mae H, Konishi T, Gao X, Tanizawa T. Semiquantitative scoring in children under treatment for atopic dermatitis. Pediatrics international : official journal of the Japan Pediatric Society 2004;46:330-6.

17. Housman TS, Patel MJ, Camacho F, Feldman SR, Fleischer AB, Jr., Balkrishnan R. Use of the Self-Administered Eczema Area and Severity Index by parent caregivers: results of a validation study. The British journal of dermatology 2002;147:1192-8.

18. Costa C, Rilliet A, Nicolet M, Saurat JH. Scoring atopic dermatitis: the simpler the better? Acta dermato-venereologica 1989;69:41-5.

19. Charman D, Varigos G, Horne DJ, Oberklaid F. The development of a practical and reliable assessment measure for atopic dermatitis (ADAM). Journal of outcome measurement 1999;3:21-34.

20. Carel K, Bratton DL, Miyazawa N, et al. The Atopic Dermatitis Quickscore (ADQ): validation of a new parent-administered atopic dermatitis scoring tool. Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology 2008;101:500-7.

21. Schmitt J, Langan S, Deckert S, et al. Assessment of clinical signs of atopic dermatitis: a systematic review and recommendation. The Journal of allergy and clinical immunology 2013;132:1337-47.

22. Wolkerstorfer A, de Waard van der Spek FB, Glazenburg EJ, Mulder PG, Oranje AP. Scoring the severity of atopic dermatitis: three item severity score as a rough system for daily practice and as a pre-screening tool for studies. Acta dermato-venereologica 1999;79:356-9.

23. JF S, A T. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. Dermatology (Basel, Switzerland) 1993;186:23-31.

24. Barbier N, Paul C, Luger T, et al. Validation of the Eczema Area and Severity Index for atopic dermatitis in a cohort of 1550 patients from the pimecrolimus cream 1% randomized controlled clinical trials programme. The British journal of dermatology 2004;150:96-102.

25. Berth-Jones J. Six area, six sign atopic dermatitis (SASSAD) severity score: a simple system for monitoring disease activity in atopic dermatitis. The British journal of dermatology 1996;135 Suppl 48:25-30.

26. Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. Archives of dermatology 2004;140:1513-9.

27. Schram ME, Spuls PI, Leeflang MM, Lindeboom R, Bos JD, Schmitt J. EASI, (objective) SCORAD and POEM for atopic eczema: responsiveness and minimal clinically important difference. Allergy 2012;67:99-106.

28. Bahmer FA. ADASI score: atopic dermatitis area and severity index. Acta dermato-venereologica Supplementum 1992;176:32-3.

29. Kunz B, Oranje AP, Labreze L, Stalder JF, Ring J, Taieb A. Clinical validation and guidelines for the SCORAD index: consensus report of the European Task Force on Atopic Dermatitis. Dermatology (Basel, Switzerland) 1997;195:10-9.

30. Rullo VE, Segato A, Kirsh A, Sole D. Severity scoring of atopic dermatitis: a comparison of two scoring systems. Allergologia et immunopathologia 2008;36:205-11.

31. Hanifin JM, Thurston M, Omoto M, Cherill R, Tofte SJ, Graeber M. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. Experimental dermatology 2001;10:11-8.

32. Schmitt J, Spuls PI, Thomas KS, et al. The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials. The Journal of allergy and clinical immunology 2014;134:800-7.

33. Chalmers JR, Simpson E, Apfelbacher CJ, et al. Report from the fourth international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME initiative). The British journal of dermatology 2016;175:69-79.

34. Torrelo A. Atopic dermatitis in different skin types. What is to know? Journal of the European Academy of Dermatology and Venereology : JEADV 2014;28 Suppl 3:2-4.

35. Margolis DJ, Gupta J, Apter AJ, et al. Filaggrin-2 variation is associated with more persistent atopic dermatitis in African American subjects. The Journal of allergy and clinical immunology 2014;133:784-9.

36. Ahmad Fadzil MH, Ihtatho D, Mohd Affandi A, Hussein SH. Objective assessment of psoriasis erythema for PASI scoring. Journal of medical engineering & technology 2009;33:516-24.

37. Hirano SA, Murray SB, Harvey VM. Reporting, representation, and subgroup analysis of race and ethnicity in published clinical trials of atopic dermatitis in the United States between 2000 and 2009. Pediatric dermatology 2012;29:749-55.

38. Bhattacharya T, Silverberg JI. Efficacy of Systemic Treatments for Atopic Dermatitis in Racial and Ethnic Minorities in the United States. JAMA dermatology 2014;150:1232-4.

39. Zhao CY, Wijayanti A, Doria MC, et al. The reliability and validity of outcome measures for atopic dermatitis in patients with pigmented skin: A grey area. International Journal of Women's Dermatology;1:150-4.

40. Charman CR, Venn AJ, Williams HC. Reliability testing of the Six Area, Six Sign Atopic Dermatitis severity score. The British journal of dermatology 2002;146:1057-60.

41. Eichenfield LF, Lucky AW, Boguniewicz M, et al. Safety and efficacy of pimecrolimus (ASM 981) cream 1% in the treatment of mild and moderate atopic dermatitis in children and adolescents. Journal of the American Academy of Dermatology 2002;46:495-504.

42. Charman CR, Venn AJ, Ravenscroft JC, Williams HC. Translating Patient-Oriented Eczema Measure (POEM) scores into clinical practice by suggesting severity strata derived using anchor-based methods. The British journal of dermatology 2013;169:1326-32.

43. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. Clinical and experimental dermatology 1994;19:210-6.

44. Chren MM, Lasek RJ, Flocke SA, Zyzanski SJ. Improved discriminative and evaluative capability of a refined version of Skindex, a quality-of-life instrument for patients with skin diseases. Archives of dermatology 1997;133:1433-40.

45. Mokkink L, Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., et al. . COSMIN checklist manual. Amsterdam, Netherlands: VU University Medical Center.; 2009.

46. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological assessment 1994;6:284.

47. Ramsay B, Lawrence CM. Measurement of involved surface area in patients with psoriasis. The British journal of dermatology 1991;124:565-70.

48. Lewis-Jones MS, Finlay AY, Dykes PJ. The Infants' Dermatitis Quality of Life Index. The British journal of dermatology 2001;144:104-10.

49. Tabolli S, Pagliarello C, Paradisi A, Cianchini G, Giannantoni P, Abeni D. Burden of disease during quiescent periods in patients with pemphigus. The British journal of dermatology 2014.

50. Albrecht J, Taylor L, Berlin JA, et al. The CLASI (Cutaneous Lupus Erythematosus Disease Area and Severity Index): an outcome instrument for cutaneous lupus erythematosus. The Journal of investigative dermatology 2005;125:889-94.

51. Rosenbach M, Murrell DF, Bystryn JC, et al. Reliability and convergent validity of two outcome instruments for pemphigus. The Journal of investigative dermatology 2009;129:2404-10.

52. Ben-Gashir MA, Hay RJ. Reliance on erythema scores may mask severe atopic dermatitis in black children compared with their white counterparts. The British journal of dermatology 2002;147:920-5.

53. Vachiramon V, Tey HL, Thompson AE, Yosipovitch G. Atopic dermatitis in African American children: addressing unmet needs of a common disease. Pediatric dermatology 2012;29:395-402.