International Journal of Women's Dermatology 1 (2015) 150–154

Contents lists available at ScienceDirect

International Journal of Women's Dermatology

The reliability and validity of outcome measures for atopic dermatitis in patients with pigmented skin: A grey area☆,☆☆

C.Y. Zhao, MBBS, MMed (Clin Epi) [a,d], A. Wijayanti [d], M.C. Doria, MD, FPDS [a,d], A.G. Harris [a,d], S.V. Jain, BSc (Med) Hons [d], K.N. Legaspi, MD, FPDS [a,d], N.C. Dlova, MBChB, FCDerm, PhD [c], M.G. Law, MA, MSc, PhD [b], D.F. Murrell, MA, BMBCh, FAAD, FACD, MD, FRCP [a,d,*]

[a] Department of Dermatology, St. George Hospital, Sydney, Australia
[b] Kirby Institute, University of New South Wales, Sydney, Australia
[c] Derpartment of Dermatology, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa
[d] Faculty of Medicine, University of New South Wales, Sydney, Australia

ARTICLE INFO

Article history:
Received 1 February 2015
Received in revised form 17 May 2015
Accepted 18 May 2015

Keywords:
Atopic dermatitis
Outcome measures
Skin of color

ABSTRACT

Background: Outcome measures for atopic dermatitis (AD) patients with pigmented skin have neither been developed nor validated.
Objective: To compare the reliability and validity of four common AD outcome measures in patients with various levels of skin darkness.
Method: The inter- and intra-rater reliability and construct validity of the EASI (Eczema Area and Severity Index), objective-SCORing Atopic Dermatitis (oSCORAD), Three Items Severity index (TIS) and Six Areas, Six Sites Atopic Dermatitis (SASSAD) were evaluated in 18 patients of various levels of skin darkness, using their full body photographs, by five trained clinicians.
Results: The inter-rater reliability intraclass coefficient (ICCs) and 95% confidence intervals were poor for highly pigmented patients: EASI -.054(-.200 to .657), oSCORAD -.089(-.206 to .598), TIS -.21(-.24 to .147), SASSAD -.071(-.200 to .631); fair for mildly pigmented patients: EASI .464(.140-.839), oSCORAD .588(.265-.89), TIS.524(.200-.865), SASSAD .41(.045-.775); and fair to good for non-pigmented patients: EASI .64(.330-.908), oSCORAD .586(.263-.889), TIS .403(.09-.809), SASSAD .667(.358-.916). Erythema likely contributed to the inter-rater variability. Construct validity had significant correlations across all measures in non-pigmented patients, but no correlations in highly pigmented patients.
Conclusion: AD outcome measures have poor reliability and validity in highly pigmented patients, with variations in erythema perception being a contributor.

Introduction

Skin color contains vital diagnostic clues in dermatology, as it reflects the underlying pathological process. Patients with pigmented skin are a complex population in dermatology. Inflammatory conditions such as atopic dermatitis (AD) and psoriasis are more difficult to assess in these patients, putting them at risk for misdiagnosis or mistreatment. Many times, patients with skin of color have been excluded from clinical trials in dermatology due to the difficulties of assessing disease severity.

There are many reasons to account for the difficulties in assessing pigmented patients with AD. Phenotypic variations secondary to genetic differences is one reason; for example, filaggrin-2 mutation variations in AD have been found in African-American patients, and have been associated with a more persistent disease course (Margolis et al., 2014; Torrelo, 2014). Another example is that pigmented skin has been shown to be less likely to develop erythema when exposed to irritants (Berardesca and Maibach, 2003). Also, cultural and environmental factors can change how the skin is cared for, leading to further heterogeneity in manifestation. In addition, the clinician's perception of color may be distorted by the background skin pigmentation or be mistaken for post-inflammatory hyperpigmentation (Ahmad Fadzil et al., 2009). Furthermore, clinical experience with managing patients with skin of color is a contributing factor. As a result, considerable intrarater and interrater variations in assessing the patient can occur, while the validity is compromised by the clinical heterogeneity.

AD is a common dermatological condition that can affect patients of all ethnicities and skin types. In the pediatric population, AD has comparably high prevalence of 17% in the United States, 14% in England, 24% in Japan, 17% in Korea, 17% in South Africa (mixed Caucasians and Blacks), 20% in Kenya, and 32% in Melbourne, Australia (Esamai et al.,

☆ Funding: This study was partly supported by the independent learning project (ILP) program and Honours program supervised by the authors M.G.L and D.F.M.
☆☆ Conflict of interest: The authors state no conflict of interest.
* Corresponding author.
E-mail address: d.murrell@unsw.edu.au (D.F. Murrell).

http://dx.doi.org/10.1016/j.ijwd.2015.05.002

2002; Oh et al., 2004; Robertson et al., 2004; Shaw et al., 2011; Simpson et al., 2009; Sugiura et al., 1998; Zar et al., 2007). A systematic review in 2012 has also found that in Africa, Eastern Asia, Western Europe and parts of Northern Europe, trends in AD were mainly increasing (Deckers et al., 2012).

The course of AD is typically chronic, requiring ongoing monitoring and an accurate assessment instrument. Also, many clinical trials have being conducted to evaluate interventions for AD, calling for a uniform outcome measure. Significant international efforts have been made to facilitate the standardization and validation of AD outcome measures. The Harmonising Core Outcome Measures (HOME) for eczema initiative has had several meetings (Chalmers et al., 2014; Schmitt and Williams, 2010; Schmitt et al., 2012), and confirmed that excoriation, erythema, edema, or papulation and lichenification are four essential components for the assessment of AD severity. A recent systematic review indicated that out of the 16 proposed outcome measures used in clinical trials, only the EASI (Eczema Area and Severity Index) and the SCORing Atopic Dermatitis (SCORAD) have received adequate validation (Schmitt et al., 2013). Meanwhile, a recent HOME consensus recommended the EASI alone as the optimal outcome measure (Chalmers et al., 2014).

Despite all of the above progress, the assessment of AD in pigmented-skin patients is still a grey area requiring attention. In fact, a study has found the underreporting of patient's skin type in clinical trials, with only 59.5% of the clinical trials published in the United States between 2000 and 2009 reporting the patient's race or ethnicity (Hirano et al., 2012). Another systematic review showed that there is a dearth of studies demonstrating efficacy of systemic AD therapy in different racial and ethnic patient subsets in the United States (Bhattacharya and Silverberg, 2014).

The aim of this study was to contribute to the expanding work in the standardization of AD outcome measures by addressing the issue of disease assessment in patients with pigmented skin. A prospective study was conducted to compare the interrater and intrarater reliability, as well as convergent construct validity, of the four most commonly used atopic dermatitis outcome measures in patients with various levels of skin darkness. This study also aimed to explore the underlying factors contributing to the variations, such as erythema.

Materials and methods

This prospective study was granted ethical approval from the South Eastern Sydney Health District Human Research Ethics Committee Northern Sector (reference: HREC/12/POWH/155).

Outcome measures tested

The outcome measures evaluated in this study were chosen as these have been most frequently validated as per a systemic review published in 2013 (Schmitt et al., 2013). These include the EASI, SCORAD, of which has a clinician-reported only version called the objective SCORAD (oSCORAD), Three Items Severity index (TIS) and Six Areas, Six Sites Atopic Dermatitis (SASSAD; Berth-Jones, 1996; Hanifin et al., 2001; Stalder and Taieb, 1993; Wolkerstorfer et al., 1999).

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 two body parts missing from their full-body photographs.

Five assessors participated in the scoring process (D.F.M, M.J.D, A.G.H, S.V.J, and K.L). All assessors were either qualified dermatologists or have been doing full-time dermatology research, and hence had been familiar with atopic dermatitis. Two (M.J.D and K.L) had trained in the Philippines and were used to darker-skinned patients. One had trained in North Carolina (D.F.M), where approximately one third of patients were African-American. All assessors were required to attend a training lecture on the use of the EASI, oSCORAD, TIS, and SASSAD. Also, the assessors were required to attend a debriefing 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 process

The assessments were performed for 2-hour sessions, over 4 days. Each session was limited to 2 hours in length, to avoid assessor fatigue. Full-body photographs of the 18 patients were presented on a screen of at least 1.5 m by 1.5 m. Three patients with various levels of skin pigmentations, whose identities were unknown to the assessors, were also arranged by a separate investigator to have their photographs repeatedly shown at the end of the 18 patients for intrarater reliability testing. For each of the patients scored, the assessors were given four color-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. 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.

Data input

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.

Categorization of skin pigmentation levels

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 categorize patients into three groups: nonpigmented (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 divide into three approximately equal categories.

Statistical analysis

All statistical analyses were performed using SPSS Version 22.0 (Armonk, NY; IBM Corp.). A professor of statistics (M.G.L) from the Kirby Institute of University of New South Wales provided help with choosing the most appropriate statistical tests.

For reliability testing, both interrater and intrarater reliabilities were assessed by the Intraclass coefficient (ICC) with 95% confidence interval (CI), using a one-way random analysis variance model. When an ICC is below .40, the clinical correlation is poor; when it is between .40 and .59, the level of correlation is fair; when it is between .60 and .74, the level of correlation is good; and when it is between .75 and 1, the level of clinical significance is excellent (Cicchetti, 1994). Scatterplots were constructed to illustrate interrater differences across all outcome measures and skin types.

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 coefficient of variations (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