Evaluation of Reliability, Validity, and Responsiveness of the CDASI and the CAT-BM

Renato Goreshi[1,2], Joyce Okawa[1,2], Matt Rose[1,2], Rui Feng[3], Lela A. Lee[4], Christopher B. Hansen[5], Carolyn A. Bangert[6], M. Kari Connolly[7], Mark D. Davis[8], Jeff P. Callen[9], Nicole M. Fett[1,2], Steven S. Fakharzadeh[1,2], Jennie T. Clarke[10] and Victoria P. Werth[1,2]

To properly evaluate therapies for cutaneous dermatomyositis (DM), it is essential to administer an outcome instrument that is reliable, valid, and responsive to clinical change, particularly when measuring disease activity. The purpose of this study was to compare two skin severity DM outcome measures, the Cutaneous Disease and Activity Severity Index (CDASI) and the Cutaneous Assessment Tool—Binary Method (CAT-BM), with the Physician Global Assessment (PGA) as the "gold standard". Ten dermatologists evaluated 14 patients with DM using the CDASI, CAT-BM, and PGA scales. Inter- and intra-rater reliability, validity, responsiveness, and completion time were compared for each outcome instrument. Responsiveness was assessed from a different study population, where one physician evaluated 35 patients with 110 visits. The CDASI was found to have a higher inter- and intra-rater reliability. Regarding construct validity, both the CDASI and the CAT-BM were significant predictors of the PGA scales. The CDASI had the best responsiveness among the three outcome instruments examined. The CDASI had a statistically longer completion time than the CAT-BM by about 1.5 minutes. The small patient population may limit the external validity of the findings observed. The CDASI is a better clinical tool to assess skin severity in DM.

Journal of Investigative Dermatology (2012) **132,** 1117–1124; doi:10.1038/jid.2011.440; published online 5 January 2012

INTRODUCTION

Dermatomyositis (DM) is a chronic systemic autoimmune disease categorized among the idiopathic inflammatory myopathies (Dugan *et al*, 2009). DM is often associated with extramuscular and extracutaneous pathology, with involvement of the joints, heart (cardiomyopathy and conduction defects), and lungs (Iorizzo and Jorizzo, 2008). The most widely accepted classification criteria for DM has traditionally emphasized the importance of clinical, laboratory, histopathological, or electrophysiological evidence of muscle inflammation for making the diagnosis (Bohan and Peter, 1975a, b). Subtypes of DM, amyopathic and hypomyopathic DM, have been described for patients with no or minor muscle findings, respectively (Gerami *et al*., 2006).

Characteristic inflammatory skin changes are seen in a large majority of individuals with DM (Callen and Wortmann, 2006). Nevertheless, the cutaneous manifestations of DM are among the least systemically studied aspects of the disease. This has resulted in part from the lack of validated tools to reliably determine the activity of the cutaneous manifestations of DM, especially relative to other dermatological diseases such as psoriasis and atopic dermatitis, where disease-specific skin severity outcome instruments have been used extensively (Kunz *et al*, 1997; Feldman and Kruger, 2005; Mrowietz *et al*., 2006; Gaines and Werth, 2008). The Federal Drug Administration has developed guidelines for researchers on how to measure clinical response through measuring disease activity, disease-induced damage, the response as determined by the patient, and health-related quality of life (Guidance for Industry Systemic Lupus Erythematosus, 2010; Gaines and Werth, 2008). From these guidelines, researchers must develop an outcome instrument that will capture appropriate elements of the disease to determine clinical response.

Currently, effective treatments for the cutaneous manifestation of DM are limited. There are a number of new biological therapies that may be beneficial for patients with DM (Iorizzo and Jorizzo, 2008). There is a critical need

[1]*Philadelphia VA Medical Center, Philadelphia, Pennsylvania, USA;* [2]*University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA;* [3]*Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania, USA;* [4]*University of Colorado, Denver, Colorado, USA;* [5]*University of Utah, Salt Lake City, Utah, USA;* [6]*University of Texas, Houston, Texas, USA;* [7]*University of California, San Francisco, San Francisco, California, USA;* [8]*Mayo Clinic, Rochester, Minnesota, USA;* [9]*University of Louisville, Louisville, Kentucky, USA and* [10]*Penn State Hershey, Hershey, Pennsylvania, USA*

Correspondence: Victoria P. Werth, Department of Dermatology, Perelman Center for Advanced Medicine, Suite 1-330A, 3400 Civic Center Boulevard Philadelphia, Pennsylvania 19104, USA. E-mail: werth@mail.med.upenn.edu

Abbreviations: CAT, Cutaneous Assessment Tool; CAT-BM, Cutaneous Assessment Tool—Binary Method; CAT-MM, Cutaneous Assessment Tool—Maximum Method; CDASI, Cutaneous Dermatomyositis Disease Area and Severity Index; DM, dermatomyositis; DSSI, Dermatomyositis Skin Severity Index; ICC, intraclass correlation coefficient; PGA, Physician Global Assessment; SRM, standardized response mean; VAS, Visual Analogue Scale

Received 14 March 2011; revised 18 June 2011; accepted 11 July 2011; published online 5 January 2012

to develop optimal validated instruments to quantify organ-specific disease activity, so that the efficacy of medications can be methodically and quantitatively evaluated.

We have previously validated a cutaneous severity outcome instrument, the Cutaneous Dermatomyositis Disease Area and Severity Index (CDASI), and have shown that it may be a more effective and reliable tool compared with other outcome measures, namely the Dermatomyositis Skin Severity Index (DSSI) and the Cutaneous Assessment Tool (CAT; Klein *et al.*, 2008). To further simplify the CDASI, we have revised the original CDASI and have shown that the modified version correlates almost perfectly with the original CDASI (Yassaee *et al.*, 2010). The CAT was originally developed with similar goals to the CDASI and was found to have appropriate reliability, construct validity, and responsiveness in the juvenile DM population (Huber *et al.*, 2007, 2008a, b). Recently, the CAT has also been simplified, and has been validated in the juvenile population (Huber *et al.*, 2008a, b). The modified versions of the CAT, named CAT-Binary Method (CAT-BM) and CAT-Maximum Method (CAT-MM), stem from an alternative scoring method of the CAT. The CAT-BM has been shown to correlate almost perfectly to the original CAT (Huber *et al.*, 2008a, b). As yet, there are no studies comparing the modified CDASI and the CAT-BM for use in longitudinal clinical research.

The current study evaluates and compares the modified tools, with a goal to provide partial validation of each tool for use in the adult DM population and to determine the optimal effective research tool for measuring the severity of cutaneous disease in adult DM. The goal is to establish an appropriate tool for evaluating DM within and between studies to evaluate therapeutic responses most effectively.

RESULTS

Distribution of scores

CDASI Total and CAT-BM Total scores had a normal distribution with scores ranging from 1 to 72 and from 1 to 20, respectively (CDASI Total: mean 24.25 ± 14.67; CAT-BM Total: mean 9.24 ± 4.17).

Inter-rater reliability

Inter-rater reliability was assessed by determining the agreement between the CDASI and the CAT-BM scores from the 10 physician raters. The CDASI was found to have good inter-rater reliability among activity and total scores and moderate inter-rater reliability in damage scores, meaning the scores among physicians were in good accordance with one another among activity and total scores and in moderate accordance with one another among damage scores. Contrastingly, the CAT-BM was found to have moderate inter-rater reliability in activity scores and poor inter-rater reliability among damage and total scores. The CDASI had the best inter-rater reliability overall when compared with the CAT-BM and PGA scales (Activity: CDASI 0.748, CAT-BM 0.516, PGA Activity 0.721, PGA Activity Likert 0.653; Damage: CDASI 0.563, CAT-BM 0.340, PGA Damage 0.506, PGA Damage Likert 0.542; Total CDASI 0.726, CAT-BM 0.432, PGA Overall 0.632, PGA Overall Likert 0.694; Table 1).

Table 1. Assessment of inter-rater reliability

Inter-rater reliability	ICC	95% CI
CDASI		
Activity	0.748	0.553–0.895
Damage	0.563	0.358–0.785
Total	0.726	0.527–0.883
CAT-BM		
Activity	0.516	0.318–0.751
Damage	0.34	0.172–0.602
Total	0.432	0.241–0.687
PGA—Activity	0.721	0.540–0.877
PGA—Activity Likert	0.653	0.446–0.860
PGA—Damage	0.506	0.313–0.743
PGA—Damage Likert	0.542	0.329–0.797
PGA—Overall	0.632	0.422–0.835
PGA—Overall Likert	0.694	0.486–0.889

Abbreviations: CAT-BM, Cutaneous Assessment Tool—Binary Method; CDASI, Cutaneous Dermatomyositis Disease Area and Severity Index; CI, confidence interval; ICC, intraclass correlation coefficient; PGA, Physician Global Assessment.

Intra-rater reliability

Intra-rater reliability measures the degree of agreement of multiple outcome scores performed by a single physician. It was assessed by determining the agreement between initial and repeat scores, using the intraclass correlation coefficient (ICC), for each outcome instrument, as well as determining the significance of a difference between mean initial scores and mean repeat scores for each outcome instrument. The CDASI was found to have an almost perfect intra-rater reliability between activity and total scores and good intra-rater reliability with damage scores (ICC: Activity 0.868; Damage 0.800; Total 0.903). No significant difference between mean initial and mean repeat activity, damage, and total scores was found (mean difference: Activity 0.00, $P=1.00$; Damage 0.40, $P=0.728$; Total −0.40, $P=0.541$). The CAT-BM was found to have good intra-rater reliability between activity, damage scores, and total scores (ICC: Activity 0.714; Damage 0.792; Total 0.800). No significant difference between mean initial and mean repeat activity, damage, and total scores was found (mean difference: Activity 0.2, $P=0.713$; Damage 0.35, $P=0.496$; Total −0.15, $P=0.634$). PGA scales were found to have almost perfect intra-rater reliability in all assessments except for PGA Activity Likert and PGA Damage Likert (ICC: 0.737 and 0.708, respectively). There was also a significant difference between initial and repeat mean scores for PGA Overall and