

3-15-2008

Alternative scoring of the cutaneous assessment tool in juvenile dermatomyositis: Results using abbreviated formats

Adam M. Huber

Peter A. Lachenbruch

Elizabeth M. Dugan

Frederick W. Miller

Lisa G. Rider

George Washington University

See next page for additional authors

Follow this and additional works at: http://hsrc.himmelfarb.gwu.edu/smhs_medicine_rheum_facpubs

 Part of the [Rheumatology Commons](#)

Recommended Citation

Huber, A.M., Lachebruch, P.A., Dugan, E.M., Miller, F.W. (2008). Alternative scoring of the cutaneous assessment tool in juvenile dermatomyositis: Results using abbreviated formats. *Arthritis Care and Research*, 59(3), 352-356.

This Journal Article is brought to you for free and open access by the Medicine at Health Sciences Research Commons. It has been accepted for inclusion in Rheumatology Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact hsrc@gwu.edu.

Authors

Adam M. Huber, Peter A. Lachenbruch, Elizabeth M. Dugan, Frederick W. Miller, Lisa G. Rider, and Juvenile Dermatomyositis Disease Activity Collaborative Study Group



Published in final edited form as:

Arthritis Rheum. 2008 March 15; 59(3): 352–356. doi:10.1002/art.23313.

Alternative Scoring of the Cutaneous Assessment Tool (CAT) in Juvenile Dermatomyositis: Results Using an Abbreviated Format

Adam M. Huber, MSc, MD¹, Peter A. Lachenbruch, PhD², Elizabeth M. Dugan, MD³, Frederick W. Miller, MD, PhD⁴, Lisa G. Rider, MD⁴, and for the Juvenile Dermatomyositis Disease Activity Collaborative Study Group

¹ *IWK Health Centre and Dalhousie University, Halifax, Nova Scotia, Canada*

² *Center for Biologics Evaluation and Research, Food and Drug Administration, DHHS, Rockville, MD**

³ *Washington Hospital Center, Washington, DC*

⁴ *Environmental Autoimmunity Group, National Institute of Environmental Health Sciences, NIH, DHHS, Bethesda, MD*

Abstract

Objectives—The Cutaneous Assessment Tool (CAT) is a comprehensive, semi-quantitative tool for the assessment of skin disease in juvenile dermatomyositis (DM). The goal of this study was to determine if alternative scoring methods would shorten the CAT, and potentially its completion time, without compromising its measurement characteristics.

Methods—One hundred and thirteen children with juvenile DM were assessed at baseline; 94 were assessed again 7–9 months later. Inter-rater reliability, internal consistency, construct validity and responsiveness were obtained using the Original scoring method and 2 alternative methods: the Maximum and Binary scoring methods.

Results—Spearman's correlations of the Maximum and Binary methods with the Original method were both 0.98 ($P < 0.0001$) for the CAT activity score and 0.96 and 0.98 respectively ($P < 0.0001$) for the CAT damage score. Values obtained for inter-rater reliability, internal consistency, construct validity and responsiveness were similar for all 3 scoring methods of the CAT. Although there was a trend towards the Maximum method having a higher inter-rater reliability, and the Binary method having a higher responsiveness, the confidence intervals were widely overlapping, and no statistically significant differences were observed. Correlations of the 3 scoring methods with other measures of skin disease activity and damage used to assess construct validity were virtually identical.

Conclusions—The Maximum and Binary methods of scoring the CAT have measurement characteristics similar to the Original method, while potentially reducing the time to administer the tool. Adoption of one of these scoring methods should increase acceptability to clinicians and researchers.

Muscle disease, expressed as weakness, poor endurance and impaired function, is the dominant feature of juvenile dermatomyositis (DM). However cutaneous involvement is also an important manifestation of both disease activity and disease damage, is an important source of morbidity, and is associated with poorer outcomes [1,2]. For these reasons, skin disease

Corresponding Author: Adam M. Huber, Division of Pediatric Rheumatology, IWK Health Centre, 5850 University Avenue, Halifax, Nova Scotia, Canada B3K 6R8, Phone: (902) 470-8827, FAX: (902) 470-7217, Email: adam.huber@iwk.nshealth.ca.

*current address for Peter A. Lachenbruch is Department of Public Health, Oregon State University, Corvallis, OR.

No reprints will be available.

activity and damage are essential components of the overall assessment of children with juvenile DM.

We have previously described the Cutaneous Assessment Tool (CAT), a comprehensive, semi-quantitative tool for assessing skin disease activity and damage in children with juvenile idiopathic inflammatory myopathy [3]. We have shown that the CAT has appropriate reliability, construct validity and sensitivity to change, and is a promising tool for use in clinical and research contexts [3,4].

In the development of the CAT, the investigators had the explicit goal of including all of the skin lesions that were important in the assessment of juvenile DM. Detailed descriptions were included for each lesion and each lesion was further sub-divided by characteristics, such as erythema, scaling or ulceration, in order to generate different levels of severity of skin disease activity or damage for each one. Individual lesions and their gradations of severity were assigned weights based on expert opinion of the importance of those features in the assessment of skin disease activity and damage. Although this resulted in a tool that was 8 pages long and quite complex, it was felt that this was necessary to reflect the complexity of the construct being measured. However, the CAT has been criticized for this length and complexity, leading to reconsideration of trying to simplify the tool.

The goal of this study was to evaluate the use of alternative scoring methods for the CAT. Specifically, we were interested in determining if alternative scoring methods would shorten the CAT, and potentially its completion time, without compromising its measurement characteristics.

Patients and Methods

Patients

This study involved reanalysis of data previously obtained [3,4], and did not involve the study of new patients. One hundred and thirteen children with definite or probable juvenile DM [5] were enrolled in this study. They were seen by 11 assessors at 10 tertiary-care pediatric institutions at baseline. Ninety-four were assessed again 7 – 9 months later. At one center, 17 children were seen by both a pediatric rheumatologist and a dermatologist within 48 hours of each other. Approval from local institutional review boards was obtained at each center, and consent was obtained from parents or legal guardians of all participants.

The characteristics of this cohort have previously been described [6–8]. In brief, children were enrolled consecutively at participating centers at any point in their disease course. At the baseline visit, the median disease duration was 19 months [25th % 8 months, 75th % 33 months] and median global disease activity and damage were 2.1 cm [25th % 0.6 cm, 75th % 4.3 cm] and 1.2 cm [25th % 0 cm, 75th % 1.5 cm] on 10 cm visual analogue scales (VAS), respectively.

All study participants were evaluated as previously described [3,4]. These assessments included history, physical examination, physician global assessments of disease and skin disease activity and damage (separate 10 cm VAS for all global assessments) [6], the Childhood Myositis Assessment Scale (CMAS) [8,9], the Childhood Health Assessment Questionnaire (CHAQ) [7], and Manual Muscle Testing (MMT) [7].

The CAT

In addition to the above, each study participant was evaluated with the CAT. Briefly, the CAT is 21-item tool with 10 items assessing skin disease activity, 4 items assessing skin disease damage and 7 items which separately assess both skin disease activity and damage, as previously reported [3]. Each item of the CAT has a specific definition and description of the

characteristics to be assessed. There are 2 – 7 response categories for each item, representing different degrees of skin disease activity or damage, to which a priori weighted scores are assigned. The scores for individual activity and damage items are summed to give the CAT activity score (potential range 0 – 96) and the CAT damage score (potential range 0 – 20). Higher scores correspond to greater activity and damage.

Alternative Scoring of the CAT

The scoring method described above, the Original Method, was compared to 2 alternative scoring methods. In the first, called the Maximum method, all items were assessed as present or absent. If an item was present, it received the maximum possible score for that item under the Original method. The activity and damage items were summed as in the Original method to give the Maximum CAT activity (potential range 0 – 96) and damage (potential range 0 – 20) scores. In the second alternative scoring method, called the Binary method, all items were also assessed as being present (scored 1) or absent (scored 0). The activity and damage items were summed to give the Binary CAT activity (potential range 0 – 17) and damage (potential range 0 – 11) scores.

Analysis

All analyses were performed using the statistical programs SAS (Release 8.02, SAS Institute Inc, Cary NC) and Stata (Intercooled Stata 9.2 for Windows, StataCorp, College Station, TX). Descriptive statistics were used including medians and Spearman's correlation coefficient where appropriate. Measurement characteristics for the Original method of scoring the CAT have been reported previously [3,4].

In order to assess inter-rater reliability of the 3 scoring methods, the intra-class correlation coefficients (ICC) were calculated for the CAT activity and damage scores for the Maximum and Binary methods and compared to the values for the Original method [3]. Confidence intervals were calculated [10].

Internal consistency was assessed with Cronbach's alpha, calculated for the total scores, and with each item deleted serially from the total score. Cronbach's alpha was calculated using standardized variables. Confidence intervals were calculated using the BOOTSTRAP function of Stata, with 1000 iterations to determine each value.

Construct validity of the alternative CAT scoring methods was assessed by calculating Spearman's correlations between the CAT activity scores obtained from the 3 scoring methods and other measures of disease activity (physician global disease activity, physician global skin disease activity, the CMAS, the CHAQ and MMT). This was repeated for the CAT damage scores obtained from the 3 scoring methods and other measures of disease damage (physician global disease damage, physician global skin disease damage, the CMAS, the CHAQ and MMT).

Responsiveness was assessed by calculating the standardized response mean (SRM) [11] for all children with data at two assessments (N = 94). The SRM for the Maximum and Binary method CAT activity and damage scores were compared to values for the Original method. Confidence intervals were calculated [11].

Power

A post hoc power analysis was conducted using the values for the original CAT activity score in order to explore the ability of this study to detect differences in some of the measurement characteristics assessed. For the ICC's, there was an estimated power of 52% to detect a

difference of 0.1 and 55% to detect a difference of 0.2. For Cronbach's alpha, there was an estimated power of 63% to detect a difference of 0.10 and 75% to detect a difference of 0.20.

Results

Baseline CAT activity and damage scores based on the Original, Maximum and Binary methods are presented in Table 1. The Spearman's correlations of the Maximum and Binary method CAT activity scores with the Original CAT activity score were both 0.98 ($P < 0.0001$). The Spearman's correlations of the Maximum and Binary method CAT damage score with the Original CAT damage score were 0.96 and 0.98 respectively ($P < 0.0001$).

Inter-rater reliability

As shown in Table 1, the ICC's for the Original, Maximum and Binary CAT activity scores ranged from 0.60 – 0.70, and from 0.65 – 0.80 for the Original, Maximum and Binary CAT damage scores. The 95% confidence intervals were widely overlapping.

Internal consistency

The Cronbach's alpha values for Original, Maximum and Binary CAT activity scores and the Original, Maximum and Binary CAT damage scores are summarized in Table 1.

Construct validity

Spearman's correlations used to compare construct validity of the 3 scoring methods are shown in Table 2.

Responsiveness

The SRM for the Original, Maximum and Binary CAT activity scores and the Original, Maximum and Binary CAT damage scores are summarized in Table 1. The values for the CAT damage score are expected to be low due the time course of the study.

Discussion

In this paper we have compared 3 methods of scoring the CAT. We have shown that the alternative scoring methods have similar measurement characteristics to the Original scoring method, including inter-rater reliability, internal consistency, construct validity and responsiveness.

For intra-rater reliability, both alternative scoring methods appeared to have somewhat lower ICC's for the CAT activity score and a lower ICC for the Binary CAT damage score. However, the 95% confidence intervals for the ICC's of all 3 methods were widely overlapping, and no significant difference in inter-rater reliability was observed. Internal consistency, as measured by Cronbach's alpha, showed a trend towards the alternative scoring methods having slightly lower values. Again, the 95% confidence intervals were widely overlapping, and no significant difference was observed. The correlations of the CAT activity and damage scores with other measures of disease activity or damage were essentially indistinguishable for the different scoring methods. Finally, the SRMs calculated to assess responsiveness were similar for each of the scoring methods. There was a trend for the SRMs of the Binary method to have somewhat higher values, but with the widely overlapping confidence intervals, no significant difference was observed.

Taken together these results fail to document superiority of any of these scoring methods. This is not surprising, given that it has been shown that changes in weighting of items do not typically affect the measurement characteristics of tools like the CAT [12]. Thus we conclude that each

of these scoring methods is likely to perform similarly. However, both alternative scoring methods simplify the CAT, and it is reasonable to adopt one of these. There is not a clear choice as to which alternative to choose. The Binary scoring method showed a trend towards better responsiveness, while the Maximum method showed a trend towards better inter-rater reliability. The Binary method results in the simplest CAT, where the assessor does not need to remember what score a particular lesion receives, while the Maximum method maintains the weighting of the Original method, and may appear to assessors to be more sensible, as it assigns larger scores for more important lesions, such as skin ulcers. The revised tool and alternative scoring methods can be found at [insert website here].

Despite their similarities in performance in this study, the alternative scoring methods have some practical advantages over the Original scoring method. The CAT has been criticized for being overly long and complex. We have argued that this allows the CAT to be more comprehensive and therefore to capture the full range of cutaneous disease in juvenile myositis [3]. Given that our results here show that the alternative scoring methods perform similarly to the Original method, maintenance of this complexity may be unnecessary. Both of the alternative scoring methods convert the CAT to a simple 21-item scale, and eliminate the gradations of severity originally described for each lesion. In doing so, the overall tool is shortened and simplified, and can now be represented on a single page. This should increase the acceptability of the CAT, likely decreasing administration time, and allowing it to be used more widely in clinical and research contexts.

There is a potential limitation to the methods that we have used in this work. It is possible that the 3 methods of scoring the CAT would have performed differently if they had actually been administered to individual children with juvenile DM. For example, it is possible that a child with very mild features of a particular lesion would be more likely to be scored as “absent” using the Maximum method than with the Binary method because an assessor might be reluctant to assign the “highest score” to a mild rash. We cannot determine if this was the case, as this study only involved rescored data that had previously been obtained.

In conclusion, we have compared 3 methods of scoring the CAT, and shown that these methods perform similarly. Given that the alternative scoring methods allow the CAT to be shortened and thus become more practical and more likely to be accepted by clinicians and researchers, we recommend that an abbreviated form of the CAT be adopted. Future work will be needed to ensure that the CAT continues to perform well in its new format.

Acknowledgements

Sources of support for this work.

This research was supported by the intramural research programs of the National Institute of Environmental Health Sciences and the National Institute of Arthritis and Skin and Musculoskeletal Diseases, National Institutes of Health, DHHS, Bethesda, MD.

References

1. Hundley JL, Carroll CL, Lang W, Snively B, Yosipovitch G, Feldman SR, et al. Cutaneous symptoms of dermatomyositis significantly impact patients' quality of life. *J Am Acad Dermatol* 2006;54:217–20. [PubMed: 16443050]
2. Huber AM, Lang B, LeBlanc CM, Birdi N, Bolaria R, Malleson P, et al. Medium- and Long-Term Functional Outcomes in a Multicenter Cohort of Children with Juvenile Dermatomyositis. *Arthritis Rheum* 2001;43(3):541–9. [PubMed: 10728746]
3. Huber AM, Dugan EM, Lachenbruch PA, Feldman BM, Perez MD, Zemel LS, et al. The Cutaneous Assessment Tool (CAT): Development and Reliability in Juvenile Idiopathic Inflammatory Myopathy. *Rheumatology*. 2007;in press

4. Huber AM, Dugan EM, Lachenbruch PA, Feldman BM, Perez MD, Zemel LS, et al. Preliminary Validation and Clinical Meaning of the Cutaneous Assessment Tool (CAT) in Juvenile Dermatomyositis. *Arthritis Rheum.* 2007;in press
5. Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). *NEJM* 1975;292(7):344–7. [PubMed: 1090839]
6. Rider LG, Feldman BM, Perez MD, Rennebohm RM, Lindsley CB, Zemel LS, et al. Development of validated disease activity and damage indices for the juvenile idiopathic inflammatory myopathies. I Physician, Parent and Patient Global Assessments. *Arthritis Rheum* 1997;40(11):1976–83. [PubMed: 9365086]
7. Huber AM, Hicks JE, Lachenbruch PA, Perez MD, Zemel LS, Rennebohm RM, et al. Validation of the Childhood Health Assessment Questionnaire in the juvenile idiopathic inflammatory myopathies. *J Rheumatol* 2001;28(5):1106–11. [PubMed: 11361197]
8. Huber AM, Feldman BM, Rennebohm RM, Hicks JE, Lindsley CB, Perez MD, et al. Validation and Clinical Significance of the Childhood Myositis Assessment Scale (CMAS) For Assessment of Muscle Function in the Juvenile Idiopathic Inflammatory Myopathies. *Arthritis Rheum* 2004;50(5):1595–603. [PubMed: 15146430]
9. Lovell DJ, Lindsley CB, Rennebohm RM, Ballinger SH, Bowyer SL, Giannini EH, et al. Development of disease activity and damage indices for the juvenile idiopathic inflammatory myopathies. II The Childhood Myositis Assessment Scale (CMAS):A quantitative tool for the evaluation of muscle function. *Arthritis Rheum* 1999;42(10):2213–9. [PubMed: 10524696]
10. Fleiss JL, Shrout PE. Approximate interval estimation for certain intraclass correlation coefficient. *Psychometrika* 1978;43(2):259–62.
11. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: Reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol* 1997;50(1):79–93. [PubMed: 9048693]
12. Streiner, DL.; Norman, GR. Health measurement scales: A practical guide to their development and use. 3. New York: Oxford University Press; 2003.

Table 1

Summary of baseline scores, inter-rater reliability, internal consistency and responsiveness for Original, Maximum and Binary methods of scoring the CAT activity and damage scores.

	Original Method	Maximum Method	Binary Method
Baseline CAT activity score			
Median	7	14	4
25 th , 75 th percentile	3, 11	6, 20	2, 5
Minimum, maximum	0, 44	0, 69	0, 12
Baseline CAT damage score			
Median	1	1	1
25 th , 75 th percentile	0, 1	0, 2	0, 1
Minimum, maximum	0, 13	0, 15	0, 8
Inter-rater reliability-CAT activity score			
Intraclass correlation coefficient	0.70	0.62	0.60
95% confidence interval	0.36, 0.86	0.28, 0.82	0.06, 0.83
Inter-rater reliability-CAT damage score			
Intraclass correlation coefficient	0.80	0.83	0.65
95% confidence interval	0.52, 0.92	0.65, 0.92	0.17, 0.85
Internal consistency-CAT activity score			
Cronbach's alpha	0.79	0.76	0.76
95% confidence interval	0.74, 0.85	0.69, 0.82	0.69, 0.82
Minimum, maximum (serial item deletion)*	0.76, 0.81	0.73, 0.77	0.73, 0.77
Internal consistency-CAT damage score			
Cronbach's alpha	0.74	0.70	0.70
95% confidence interval	0.42, 1.0	0.41, 0.99	0.41, 0.99
Minimum, maximum (serial item deletion)*	0.67, 0.76	0.61, 0.73	0.61, 0.73
Responsiveness-CAT activity score			
Standardized response mean	0.52	0.49	0.67
95% confidence interval	0.32, 0.72	0.29, 0.69	0.47, 0.87
Responsiveness-CAT damage score			
Standardized response mean	0.02	0.02	0.16
95% confidence interval	-0.18, 0.22	-0.18, 0.22	-0.04, 0.36

* This is the minimum and maximum values of Cronbach's alpha that were obtained when each item was serially deleted from the calculation of the total score.

Table 2**Summary of the comparison of the construct validity of the 3 methods of scoring the Cutaneous Assessment Tool (CAT)**

Construct validity was assessed using Spearman's rank correlations of the CAT activity and damage scores with other measures of disease activity and damage. All $P < 0.0001$ except as noted.

	N	CAT activity score		
		OSM	MSM	BSM
Physician Global Assessment of Skin Disease Activity	96	0.83	0.81	0.81
Physician Global Assessment of Disease Activity	98	0.77	0.75	0.76
Childhood Myositis Assessment Scale	98	-0.48	-0.46	-0.47
Childhood Health Assessment Scale	96	0.40	0.40	0.42
Total Manual Muscle Testing Score	50	-0.36*	-0.39*	-0.40*
		CAT damage score		
		OSM	MSM	BSM
Physician Global Assessment of Skin Disease Damage	95	0.53	0.54	0.49
Physician Global Assessment of Disease Damage	113	0.52	0.51	0.47
Childhood Myositis Assessment Scale	98	-0.19 [†]	-0.18 [†]	-0.18 [†]
Childhood Health Assessment Scale	96	0.13 [†]	0.15 [†]	0.11 [†]
Total Manual Muscle Testing Score	50	-0.38*	-0.34*	-0.37*

* $P \geq 0.0001$ and < 0.05

[†] $P \geq 0.05$

Abbreviations: OSM=Original scoring method, MSM=Maximum scoring method, BSM=Binary scoring method