Quantifying Skin Disease Burden in Mycosis Fungoides–Type Cutaneous T-Cell Lymphomas

The Severity-Weighted Assessment Tool (SWAT)

Seth R. Stevens, MD; Malcolm S. Ke, MD; Eileen J. Parry, MD; Julie Mark, MD; Kevin D. Cooper, MD

Objective: To develop a quantitative tool to assess severity of mycosis fungoides.

Design: Prospective analysis of a cohort.

Setting: University department of dermatology–based cutaneous lymphoma clinic.

Patients: From 1984 to 1995, 1186 visits from 323 referred patients seen in a multidisciplinary cutaneous lymphoma program.

Main Outcome Measures: Severity-weighted assessment tool (SWAT) scores were obtained for patients at each visit. This score represents the product of the percentage total body surface area (%TBSA) involvement of each lesion type (patch, plaque, and tumor or ulceration), multiplied by a weighting factor: SWAT = (patch %TBSA × 1) + (plaque %TBSA × 2) + (tumor or ulcer %TBSA × 3). In addition, the standard measurements of TBSA involvement and physician global assessments were recorded for comparison.

Results: The SWAT score correlated well with %TBSA (r = 0.95, P < .001), physician global assessment (r = 0.60, P < .001), and time to complete remission during psoralen–UV-A therapy (r = 0.80, P < .001), therefore indicating validity against standard measures. Analysis of individual and subsets of patients demonstrated that the SWAT score more accurately quantified changes in skin disease burden, including mixed responses to treatment, than did %TBSA alone.

Conclusions: The SWAT score is a useful clinical measurement for mycosis fungoides. The SWAT score captures overall physician impressions of disease status on a continuous dimensionless numerical scale, therefore providing a defined, objective, and sensitive quantitative measure. This tool is suitable for individual patient assessment, clinical trials, and outcome comparisons.

Arch Dermatol. 2002;138:42-48

THE QUANTIFICATION of disease burden is necessary for the accurate documentation of disease status and is essential to the assessment of response to therapeutic interventions in clinical practice and research. Several assessment tools exist for the quantification of inflammatory dermatologic disease, including the Psoriasis Area and Severity Index, Severity Scoring of Atopic Dermatitis, and Atopic Dermatitis and Severity Index. The ability to assess disease by area of involvement is one of the most important aspects of these measures, but objectiveness is lacking in all but the Atopic Dermatitis and Severity Index. The other methods assess percentage total body surface area (%TBSA) in a more subjective manner based on “the rule of 9s,” visual estimates, comparison with body sites of known area (e.g., the palm is 1%), or a combination of these. Estimates of area of involvement by these means are inaccurate; fewer than 50% of observers estimated within 25% of the true area in one study using schematic figure outlines. In contrast, the Atopic Dermatitis and Severity Index uses an objective method to determine the area of skin involvement by mapping disease onto body diagrams and then evaluating the extent by grid-point counting. This is a simple method of area estimation that uses a regularly spaced square grid placed over the body diagram of interest. The area fraction of any skin severity grade drawn on the diagram is estimated by counting the number of point intersections that fall on the body diagram and then evaluating the extent by grid-point counting. This is a simple method of area estimation that uses a regularly spaced square grid placed over the body diagram of interest. The area fraction of any skin severity grade drawn on the diagram is estimated by counting the number of point intersections that fall on the body diagram and then evaluating the extent by grid-point counting.
PATIENTS AND METHODS

PATIENTS

Data were available for all patients seen in a multidisciplinary cutaneous lymphoma program from 1984 to 1995, comprising 1194 data records analyzed from 323 patients. Records of 6 evaluations containing SWAT scores that were numerically invalid and 2 evaluations with incomplete SWAT data were excluded, leaving 1186 records for further consideration. Records of 21 evaluations lacked physician global assessment (PGA) data and were excluded from analyses involving this factor.

ASSESSMENTS

The diagnosis of cutaneous T-cell lymphoma was made on the basis of histologic, molecular diagnostic, and clinical criteria.12 At each visit, a PGA was performed before calculating the SWAT score. A PGA was rated as clear of disease (rating 0), mild (rating 1), moderate (rating 2), or severe (rating 3), a measure that integrates the physician’s and patient’s evaluation of MF status. Next, SWAT forms (Figure 1) were completed. Clinical disease was mapped onto body diagrams by 2 dermatologists (S.R.S. and K.D.C.). Patch disease was defined as flat erythema and was represented by single-hatched markings, while plaque disease was defined as an elevated area and was represented by crosshatching. Tumors were defined as dome-shaped, nodular lesions of greater than 1-cm elevation, while ulcerative lesions were those with significant loss of superficial skin, including the entire epidermis and some portion of the upper dermis. These were represented as solid shaded areas. In erythrodermic MF, the degree of edema or infiltration was used for mapping skin severity: erythroderma with mild infiltration was mapped as patch disease, erythroderma with moderate infiltration was mapped as plaques, and erythroderma with tumorous infiltration or ulceration (including fissuring) was mapped as tumors or ulceration. At follow-up visits, if lesions were flat but erythema or atrophy persisted, a biopsy specimen was taken to distinguish residual MF from residual therapy, and the lesion was outlined but left unshaded until results were available. A 1-cm grid was randomly placed on the drawing, and the number of grid intersections overlying each type of lesion was counted. The mean of 6 such grid placements and counts was divided by the maximum possible number of intersections for the body diagram and multiplied times 100 to yield the %TBSA involved with each lesion type. The sum of these 3 numbers was the extent of or %TBSA involvement, on a 0 to 100 scale. Severity weighting was achieved by multiplying the area for patches by 1, the area for plaques by 2, and the area for tumors or ulcers by 3. The sum of these 3 numbers was the dimensionless SWAT score, on a 0 to 300 scale: SWAT=(patch %TBSA×1)+(plaque %TBSA×2)+(tumor or ulcer %TBSA×3).

STATISTICAL ANALYSIS

Data were entered prospectively into a database (dBase IV; Microsoft Corp, Redmond, Wash), with ongoing physician and nursing review of printed reports of the patients’ data records. The fields containing %TBSA, SWAT score, and PGA were then transferred to a spreadsheet (Excel version 7.0, Microsoft Corp) for analysis. Data were analyzed using linear regression analysis with 95% confidence intervals. Correlations were considered significant at P<.05.

Figure 1. Severity-weighted assessment tool form for a hypothetical patient with mycosis fungoides. Patch disease is drawn as single-hatched, plaque disease as crosshatched, and tumors and ulceration as solid shaded areas. A point-counting grid is used to calculate area of involvement by each lesion type.

©2002 American Medical Association. All rights reserved.
use of vasoconstricting topical steroids. Also, psoralen-UV-A (PUVA) therapy, commonly used to treat MF, can result in persistent erythema at old lesion sites that are actually clear of disease. In addition, erythema scores that separate barely detectable, readily detectable, and marked and maximal erythema are not easily and consistently distinguishable. In contrast, observers of skin disease can readily distinguish lesion types: patch, plaque, and tumor or ulceration factors reflect the histologic depth of infiltration and are thereby valid measures of disease burden or severity. Finally, the severity of erythema, scaling, or induration may vary in intensity relative to the other two. The links between these 3 factors make accurate assessment difficult.

Others have described a tumor burden index (TBI) for MF that is based on a retrospectively derived equation, the most recent being: TBI = 1 + (patches × 2) + (plagues × 2) + (tumors × 1.3), where the factor for patches equals zero if there is 30% or less TBSA patch involvement, 1 if greater than 30% TBSA, 1 if plaques or tumors are present, and 0 if absent. Based on the most recent report, 6 possible TBI scores exist (1, 2.3, 3, 4.3, 5, and 6.3). In previous reports, varying weighting factors were used to demonstrate that the TBI correlated with the level of soluble interleukin-2 receptors, a marker for disease activity, and the development of interferon antibodies during treatment with interferon alpha-2a in combination with PUVA or retinoids. However, more recently, this group suggests the TBI to be most useful as a prognostic tool for individual patients at the time of presentation. This score could then help guide treatment strategies based on their survival expectancy. The authors further point out that the TBI cannot be used to monitor changes in a patient’s disease.

In the present report, we describe a severity-weighted assessment tool (SWAT) for skin lesions in MF, using the principle of grid-point counting to facilitate accurate assessments of area of skin involvement by each lesion type. This study validates the use of the SWAT as a sensitive and dynamic assessment tool for measuring cutaneous disease in MF. We demonstrate that the SWAT is responsive to clinically significant changes in patients’ status and more accurately reflects the nature of a patient’s skin disease than do estimates of %TBSA involvement. In contrast to the aforementioned TBI, the usefulness of which lies in disease prognosis governing treatment strategies, the SWAT is primarily useful in clinical studies to sensitively measure changes in a patient’s skin condition.

![Figure 2](https://example.com/fig2.png)

**Figure 2.** The severity-weighted assessment tool (SWAT) score correlates with the percentage total body surface area (%TBSA) of skin involvement in the assessment of patients with mycosis fungoides (r = 0.95, P < .001, n = 1186). Patient assessments in rectangles A and B are used for analysis of physician global assessments in Figure 3.

**RESULTS**

The dermatologist is accustomed to considering the %TBSA involved with disease as a measure of disease burden. Therefore, %TBSA of involvement is used as a comparative standard assessment in a scatterplot of each evaluation vs SWAT score (Figure 2). Rectangles A and B define subsets of patients with similar %TBSA or SWAT scores, which are further examined herein and in Figure 3. As shown in Figure 2, the SWAT score correlates well with %TBSA (r = 0.95, P < .001, n = 1186).

To rule out the possibility that the strong correlation of %TBSA and SWAT score is mainly driven by those points with only one type of lesion, we reanalyzed the data, excluding points along the lines with slopes of exactly 1 (all patches), 2 (all plaques), or 3 (all tumors or ulceration), including evaluations in which %TBSA and SWAT score equaled zero (no lesions). Because some patients with mixed lesions may fall exactly on the line with a slope of 2, we may have erred on the side of conservatism and excluded a few data unnecessarily. For patients with mixed lesion types, the association is in fact slightly stronger (r = 0.98, P < .001, n = 118, data not shown). Therefore, SWAT score correlates with an accepted standard assessment indicating validity, while distinguishing additional factors of severity.

**THE SWAT MORE ACCURATELY CORRELATES WITH PGA SCORES OF MF STATUS THAN DOES %TBSA ALONE**

The PGA (mild, moderate, or severe) correlated well with SWAT (r = 0.60, P < .001, n = 1165, data not shown) and %TBSA (r = 0.60, P < .001, n = 1165, data not shown). However, to determine whether the variation in SWAT scores at a given %TBSA of involvement represented a clinically significant distinction of disease severity, we analyzed subsets of patients further.

When we examined patients’ data records from a narrow midrange within 10% of %TBSA (45%-55%, Figure 2, rectangle A), all possible PGA scores were represented: 1 mild, 4 moderate, and 3 severe (Figure 3A). Therefore, a homogeneous population of %TBSA values does not have homogeneous PGA scores. More important, the variation in the PGA scores for the midrange of %TBSA values correlated with the corresponding SWAT scores (r = 0.80, P = .02, n = 8). Furthermore, the y-intercept of the linear regression model is 11.9, which
reflects that as the PGA score goes toward zero (no disease), the SWAT score also goes toward zero. Therefore, heterogeneity of PGA scores within a narrow range of %TBSA data is sensitively detected by the SWAT but not by %TBSA alone, indicating a superior ability to capture overall physician impressions of disease status in a defined, objective, quantitative manner.

In contrast, when we examined patients’ data records from a narrow midrange within 10% of SWAT scores (135-165, Figure 2, rectangle B), there was minimal variation in PGA scores, which in turn correlated poorly with the corresponding %TBSA values (Figure 3B). Furthermore, the y-intercept of the linear regression model was 97.6, implying that as the PGA scores approach zero (no disease), the %TBSA will approach total body involvement, a logically untenable prediction. Therefore, the SWAT score outperforms %TBSA, as it correlates more closely with variation in PGA. We also looked at the upper and lower quartile ranges within 10%; 22.5% to 27.5% TBSA vs 67.5 to 82.5 SWAT, and 67.5% to 82.5% TBSA vs 202.5 to 247.5 SWAT, with similar results for subjects at the midpoint of each scale (data not shown). Therefore, the PGA scores of patients with similar SWAT scores were more homogeneous than those with similar %TBSA values. Linear regression analysis also showed that SWAT correlated better with the variation in appearance of those selected by %TBSA than vice versa, with y-intercepts closer to zero (not shown). These data confirm the advantage of the SWAT over %TBSA involved as a measure of skin disease burden in MF.

THE SWAT SENSITIVELY AND RESPONSIVELY MEASURES CHANGE IN PATIENTS’ STATUS DURING THERAPY

Another measure of validity is whether an assessment tool captures change in disease burden during successful therapy. Psoralen–UV-A photochemotherapy has been shown to be an effective therapy for MF-type cutaneous T-cell lymphomas.18 In 42 patients with MF treated with PUVA, 34 achieved complete remission within 3 months, 2 achieved complete remission between months 3 and 5, and 6 did not achieve complete remission (100% clearing). We plotted time after initiating PUVA vs SWAT score normalized as percentage of baseline SWAT score from 98 evaluations of the 34 patients (98 evaluations) achieving complete remission of MF within 3 months (r=0.80, P<.001). The larger plus symbols denote all 34 patients at months 0 (100% baseline disease) and 3 (0% baseline disease) of PUVA therapy.

Figure 4. The severity-weighted assessment tool (SWAT) captures progressive improvements in mycosis fungoides (MF) during psoralen–UV-A (PUVA)–induced remissions. The SWAT correlates with time on PUVA therapy in the 34 patients (98 evaluations) achieving complete remission of MF within 3 months (r=0.80, P<.001). The larger plus symbols denote all 34 patients at months 0 (100% baseline disease) and 3 (0% baseline disease) of PUVA therapy.
scores than by %TBSA alone. Figure 5 documents clinical assessment factors in a patient with erythroderma receiving electron beam radiation followed by a maintenance dosage of topical mechlorethamine hydrochloride. Over time, all 3 disease measurement factors reach zero (no disease, data not shown). However, during the first 3 assessments, %TBSA disease remains the same at 99%, while the SWAT score drops from 297 to 190. Likewise, the PGA also documents an improvement from severe to moderate disease status as lesion thickness and induration diminish. Therefore, the SWAT provides improved sensitivity and objectivity and accommodates erythrodermic and patch- or plaque-type MF. This patient also demonstrates that PGA does not have a broad dynamic range. With only 4 possible assessments, little discrimination of disease status is possible.

THE SWAT SENSITIVELY DESCRIBES MIXED RESPONSES AND PROGRESSION

Data from a second patient with MF, treated with electron beam radiation followed by PUVA and interferon alfa-2a, are shown in Figure 6. Between assessments 1 and 3, the patient experienced a mixed response, during which time some patches reduced in size or resolved, while other areas progressed and ulcerated. The new appearance of ulcerated plaques represented significant additional morbidity for this patient, which was not compensated by clearing a few patches. This worsening disease status (from moderate to severe disease) was perceived globally by the physician and was reflected as an increase in SWAT. However, the %TBSA values do not capture the patient’s deterioration between assessments 1 and 3. Therefore, the SWAT more specifically describes the mixed response to therapy and focal but symptomatic disease progression than does %TBSA alone.

The objective quantification of disease is an important element of patient evaluation. In recent years, health care outcomes research and related cost-benefit analyses have become increasingly important. The SWAT in patients with MF is a sensitive and precise tool with a broad dynamic range. It also provides a hard copy of the primary data at the same time that it provides a single numerical summary that quantifies the results of cutaneous examinations of patients with MF. The advantages of such an instrument are that the single SWAT score can be applied to all patients with MF and provides summary data that can be conveniently analyzed. It also preserves potentially important data for subsequent analysis and can easily accommodate modifications based on new prognostic indicators as they become available. These advantages make this system particularly useful for database entries, tracking of patients’ responses to therapies, clinical trials, and outcomes research.

Criterion validity defines how well a new measure compares with the gold standard. However, in MF there is no established gold standard, with published clinical assessments in trials using scales of improvement (progression, no response, partial response, and complete remission), %TBSA, or size of target lesions. The SWAT score correlates well with %TBSA and PGA, indicating validity as a measure of disease in MF. The SWAT score also demonstrated changes over time of individual and group responses. Individual assessments that diverged from expectation when assessed by %TBSA alone were more accurately assessed by SWAT score. The increased sensitivity of the SWAT method over %TBSA to reflect changes in patients’ status, such as the resolution of plaques into patches, may be particularly important in identifying patients who are responding early during treatment before complete clinical clearing of lesions. The SWAT score also provides greater flexibility than the use.
of PGA, with 300 distinct scores for SWAT vs only 4 for PGA. With the use of this system, a patient with patch or plaque disease covering 5% TBSA might be scored as having mild disease, a score of 1. However, all patients with trivial disease, such as those almost clear during treatment or with one remaining patch, also have to be scored as having mild disease, as this is the lowest point on the assessment scale. A score of 1, which is one third from the baseline (no disease), does not reflect the actual disease present and provides little flexibility in scoring patients with minimal involvement. Therefore, SWAT has advantages over PGA.

The SWAT is particularly useful for 2 common types of assessments. First, during routine patient care, the SWAT most accurately quantifies skin lesions in MF. Therefore, patients’ progress, or lack thereof, can be followed over time, and treatment efficacy can be accurately documented. For example, it has been shown that photopheresis therapy in a patient with Sézary syndrome is much less likely to result in long-term survival (>3 years) if a 50% reduction in skin score has not been achieved within 5 months of therapy.25

The second type of assessment in which the SWAT is most valuable is for the evaluation of data collected over time from many patients. An example would be patient assessment during clinical trials. The SWAT is useful during the initial phase in which different treatment groups may need to be stratified according to skin severity and later to follow up patients’ response to intervention. In addition, documentation of patch, plaque, and tumor or ulcerated lesions in databases, which are being established at several centers, will allow for future refinement of our understanding of MF. Of particular note in this last regard would be the potential need to refine the T rating of standard MF staging. Although the SWAT is not intended as a staging factor, further development and validation will help ascertain whether it has prognostic significance. A recent evaluation of prognostic factors in MF identified that the type of skin disease at initial diagnosis was one of the best prognostic indicators of survival and clinical outcome.26 Because the SWAT precisely documents cutaneous disease burden in MF, it may be a useful factor in future assessments of prognosis and survival, compared with standard staging assessments. For example, the types of data accumulated by the SWAT will determine whether it is reasonable to combine patch and plaque disease and whether 10% skin involvement is the appropriate dividing line between T1 and T2.

Previously published methods of assessing skin disease in MF have used a method similar to the Psoriasis Area and Severity Index to estimate disease burden.23 This approach generates confusion when rating body areas with mixed lesion types. Should the lesion severity be averaged over the defined body area, or should the most severe lesion be chosen as representative? For example, it is unclear how to grade a patient who has severe disease covering 0.5% of the back and mild disease covering 30% of the back. The investigator is forced to choose between rating the back as severe disease, mild disease, or some intermediate grade. The SWAT does not force the investigator to develop intermediate grading systems on an ad hoc basis. In addition, predetermined grouping of skin signs together10 does not allow flexibility for a mixed severity of different signs. An example would be severe erythema with only mild scaling. In contrast, the SWAT method allows all areas of the body to be individually rated and relies only on differentiation of lesion type (patch, plaque, and tumor or ulceration) and degree of infiltration (if erythrodermic), rather than on the assessment of clinical signs (ie, erythema and scaling).

Although the TBI13-15 may seem similar in method to the SWAT, its derivation and application are different. As the authors point out, the TBI functions mainly as a prognostic tool, not as a descriptive indicator of disease severity. For example, a patient with 90% TBSA patch disease could not be distinguished from another patient with only 35% TBSA patch disease, as both TBIs would be 3. It was not designed to sensitively differentiate the 2 clinically different patients. The TBI also cannot dynamically reflect a change in a patient’s disease. For example, if the same patient with 90% TBSA patch disease were involved in a clinical trial and demonstrated more than 50% reduction to only 35% TBSA patch disease, he or she would be considered as having experienced a partial response, but would have a static TBI of 3. On the other hand, the SWAT score was primarily designed to sensitively and dynamically reflect these changes. This same patient would initially present with a SWAT score of 90 and then drop to 35, indicating skin improvement. In discussing the derivation of weighting factors, the most recent TBI weighting factors (2 for patches, 2 for plaques, and 1.3 for tumors) were retrospectively derived, while those for the SWAT (1 for patches, 2 for plaques, and 3 for tumors or ulcers) were prospectively determined. These values were chosen arbitrarily based on intuitive elements. For example, the presence of tumors would naturally be assigned a more severe weighting factor of 3 compared with plaques (factor 2) and patches (factor 1). The TBI weighting factors, because of their retrospective derivation, seem less intuitive. The 1.3 weighting factor of a tumor is unexpectedly less than that of plaques (factor 2) and patches (factor 2) even though it is considered more severe. This discrepancy in choosing weighting factors to accurately reflect disease burden will require further study. Also, we are not aware that the TBI has been validated against the standard methods of disease assessment in MF, while the SWAT has been shown to outperform measurement of disease by %TBSA alone and to correlate well with PGA. Moreover, the SWAT can sensitively capture changes in the disease burden of patients undergoing treatment.

The SWAT scoring system for MF does not assess all aspects of the disease. It relates to the %TBSA and nature of physical skin lesions, with no measure of internal involvement, psychosocial disability, or comorbidity. However, physical, psychological, and social aspects of health are closely related. In patients with MF, physical symptoms (pain, itch, and scaling), psychosocial disability, and overall quality-of-life issues are likely to be closely linked to the severity and extent of cutaneous disease. The SWAT method will provide an informative, continuous scale instrument by which valid comparisons with formal assessments of quality of life in MF (such as the Skindex27) can be made. The SWAT has also not been evaluated as a prognostic or staging tool, but only as an indicator of skin dis-
ease severity. The prognostic implications of the SWAT and its use in staging are also areas of future study.

The assumptions inherent in the SWAT are that assessors can consistently distinguish lesion types and can accurately draw proportionally on assessment forms. Interobserver and intraobserver variability still need to be examined. Also, the current weighting of lesion types may not be an accurate estimate of increase in disease burden. Studies are under way to determine the validity of these assumptions. The results of such further investigations will help to refine the SWAT, but several issues are still unresolved. An assessment of actual tumor burden in MF is impaired by the inability to accurately distinguish malignant vs reactive T cells within individual lesions. It remains unclear how to differentially weight lesions that have different proportions of these 2 cell types, as they have opposite implications with respect to pathomechanisms. Therefore, we specifically use the term disease burden based on the assumption that malignant and reactive T cells contribute to the disease manifestations patients have. Since the introduction of the SWAT into our routine clinical practice, we have made slight modifications in the process of calculation in that we now use a 0.5-cm grid and make only 3 random placements of the grid while point counting. The time taken to complete the SWAT form depends on the total amount of skin disease that must be proportionally drawn on the body diagram. In our experience, this usually takes less than 5 minutes in a patient with localized disease and no more than 10 minutes in a patient with extensive skin disease (data not shown). A comparison of less time-consuming methods to assess area of involvement, such as simple visual estimates, with those used in the SWAT will require further study to see if they are able to define area involved with acceptable accuracy for the measurement of disease burden.

In summary, the SWAT is a valid and sensitive dimensionless method to assess skin involvement with MF. It correlates well with other measures of disease, such as %TBSA and PGA. In situations in which the SWAT scores are at odds with %TBSA alone, the SWAT more accurately and objectively captures important clinical factors. Compared with assessments like the Psoriasis Area and Severity Index, in which one can assign only a single lesion severity to a predefined area of the body, the SWAT score allows unlimited capture of mixed lesions in all areas of the body in their actual proportions. Additional advantages of the SWAT are that it is on a continuous, numerical scale that is useful in monitoring incremental changes in an individual's disease status or in comparing results from different patients. Such sensitive measurement is particularly useful in monitoring early or partial response to treatment, such as during clinical trials of new therapies, and in the future understanding of clinical factors that may determine the outcomes of patients with MF.

Accepted for publication April 26, 2001.