Type of paper: Original research

Title: Measuring Quality of Life in Patients with Mycosis Fungoides (MF)/Sézary Syndrome (SS) Cutaneous T-cell Lymphoma

Authors:

Stacey A McCaffrey, PhD
PatientsLikeMe, Inc., Cambridge, MA, USA, smccaffrey@patientslikeme.com

Ryan A Black, PhD
College of Psychology, Nova Southeastern University, Fort Lauderdale, FL USA,
blackrya@nova.edu

Mitchell Nagao, PharmD
Actelion (US), Inc., South San Francisco, CA, USA, mitchell.nagao@actelion.com

Marjan Sepassi, PharmD
Actelion (US), Inc., South San Francisco, CA, USA, marjan.sepassi@actelion.com

Gaurav Sharma, PharmD
Actelion (US), Inc., South San Francisco, CA, USA, GSharm23@its.jnj.com

Susan Thornton, BBA
Cutaneous Lymphoma Foundation, Birmingham, MI, USA, susan@clfoundation.org

Youn H Kim, MD
Stanford University, Stanford, CA, USA, younkim@stanford.edu

Julia Braverman, PhD
PatientsLikeMe, Inc., Cambridge, MA, USA, braverju@gmail.com

Corresponding author:

Stacey McCaffrey, PhD, PatientsLikeMe, Inc., 160 Second Street, Cambridge, MA 02142, USA
Development of MF/SS-CTCL QoL

Telephone: 617-499-4003 Fax: 866-850-6240 Email: research@patientslikeme.com
Abstract

**Background:** Although quality of life (QoL) plays an important role in treatment decision-making and clinical management of Mycosis Fungoides or Sézary Syndrome subtypes of cutaneous T-cell lymphomas (MF/SS-CTCL), a MF/SS-specific measure of QoL does not exist.

**Objective:** The objective of this research was to develop and validate the first QoL instrument for MF/SS-CTCL using a patient-centered approach.

**Methods:** A conceptual framework for the MF/SS-CTCL QOL was developed through a literature review and interviews with key opinion leaders. Concept elicitation with patients was utilized to refine the conceptual model and generate preliminary items. The items were then revised based on qualitative and quantitative feedback obtained through cognitive debriefing surveys and interviews with patients. Next, participants (N=126) completed the preliminary MF/SS-CTCL QOL and a comparator measure of health-related quality of life (Skindex-29) through the PatientsLikeMe Open Research Exchange. Sixty-six participants completed the MF/SS-CTCL QOL again five days later for purposes of evaluating test-retest reliability. The MF/SS-CTCL QOL was finalized based on results from an empirical evaluation, which included both classical and modern test theory approaches. Specifically, this included evaluation of: (1) the optimal item response theory (IRT) measurement model; (2) item fit; (3) unidimensionality; (4) rating scale performance; (5) reliability; (6) test information (precision); (7) person-to-item map; (8) convergent and discriminant validity; and (9) presence of bias via differential item function.

**Results:** Results from the comprehensive psychometric evaluation utilizing a Rasch-Grouped Rating Scale yielded a final 12-item instrument. The rating scale was functioning as expected, and the instrument exhibited adequate person reliability (.87), good to excellent test-retest...
reliability ($r=.89, P < .001$), high levels of measurement precision, and good person-to-item targeting. The correlation between the MF/SS-CTCL QoL and the Skindex-29 ($r=.852, P < .001$) was significantly greater than the correlation between the MF/SS-CTCL QoL and syndrome stage ($r=.260, P < .001$), providing support for convergent and discriminant validity. Items did not evidence significant bias based on gender, age, or race. Rasch scores were converted to scaled scores with qualitative descriptive categories for ease of interpretation.

**Conclusions:** Empirical evaluation demonstrated strong evidence of excellent psychometric properties. Utilizing a patient-centered measure development approach ensures that this QoL instrument captures the information that is most meaningful and clinically relevant to patients.

**Keywords:** quality of life, Rasch, patient-reported outcome, cutaneous lymphoma, Mycosis Fungoide, Sézary Syndrome
Introduction

Mycosis fungoides (MF) and its leukemic variant Sézary syndrome (SS) represent approximately 65% of the cases of cutaneous T-cell lymphoma (CTCL), a class of Non-Hodgkin’s lymphomas with a relapsing course over the span of decades [1] [2]. For patients with MF/SS-CTCL, quality of life (QoL) plays an important role in treatment decision-making and clinical management of the disease. Currently, several cancer-specific (EORTC tools [3], FACT-G [4]) and skin-specific (Itchy-QOL [5], DLQI[6], Skindex-29 [7], VAS itch[3]) health-related quality of life instruments exist. Although clinicians often use these instruments or a combination of these instruments to estimate QoL for patients with MF/SS subtypes of CTCL (MF/SS-CTCL), this strategy can be time consuming and burdensome for patients. Additionally, these patient-reported outcome (PRO) instruments were not specifically designed to capture the unique experiences of patients living with MF/SS-CTCL and may not accurately quantify QoL for this patient population. Further, clinical assessments of MF/SS-CTCL (e.g., extent of the lesion) may not be an appropriate proxy to capture the impact of the disease on QoL [8]. Therefore, a patient-centered disease-specific PRO to measure QoL for patients with MF/SS-CTCL is urgently needed to improve quality of care for these patients and to progress research within this clinical arena. The purpose of the present study is to fill this critical gap by developing and validating the first QoL instrument specifically developed by and designed for patients with MF/SS-CTCL—the “MF/SS-CTCL QoL.”

The MF/SS-CTCL QoL was developed in two broad phases; (1) instrument development and (2) psychometric evaluation. Methods and results of each phase are described separately below.
Methods

Instrument Development

The purpose of instrument development was to create items that comprehensively capture the different facets of QoL that are impacted by MF/SS-CTCL. Patients were closely involved in the item development process to ensure that the final instrument evaluated aspects of QoL that were most relevant and meaningful for them. Instrument development involved three primary steps: (1) creating a conceptual framework, (2) concept elicitation, and (3) cognitive debriefing. This research was approved by the New England Institutional Review Board.

Creating the Conceptual Framework

The conceptual framework of QoL for MF/SS-CTCL patients was developed through a literature review and interviews with key opinion leaders. Physicians and experts in the field of cutaneous lymphomas (N=3), participated in interviews to gather information related to treatment, challenges in caring for and treating patients with MF/SS-CTCL, main concerns expressed by patients, impact of the condition on patients’ well-being and daily functioning, use and availability of PRO instruments, and unmet needs within the research and patient care field. Results from the literature review and key opinion leader interviews highlighted the importance of evaluating condition-specific facets of QoL, such as physical functioning, emotional functioning, and social functioning. Key opinion leaders also indicated that two facets of QoL—coping and self-management—were absent from existing PRO measures and may be important for patients with MF/SS-CTCL.

Concept Elicitation

The purpose of concept elicitation was to gather patient feedback regarding their experience of living with MF/SS-CTCL and generate preliminary items. Data for concept elicitation was
Development of MF/SS-CTCL QoL

collected from patients through a survey conducted through the PatientsLikeMe online research
platform (Open Research Exchange [ORE]) and follow-up interviews conducted using
phone/video conferencing. Patients were eligible to participate if they were members of
PatientsLikeMe, adults, and reported a diagnosis of MF or SS. Survey content included
demographic and clinical items, as well as open-ended questions pertaining to health-related QoL
derived from the conceptual model. Follow-up interviews consisted of semi-structured questions
based on the participant’s responses to the survey.
Two trained raters coded the data independently using MAXQDA® software. The codebook was
finalized after the satisfactory inter-rater agreement (Cohen’s Kappa of 0.65 or greater) was
reached. Content saturation was assessed across patients with a saturation table where saturation
was reached when no new information was obtained through data collection [9]. The codes with
best agreement and highest frequencies were selected and then grouped by themes to generate
the initial items for the QoL instrument.

Results from Concept Elicitation

Twenty-one participants completed the online survey, and 10 of those participants completed a
follow-up interview. Sixty-seven percent of the sample was female, all were white, non-
Hispanic, and the average age was 55 years old (SD =12.39). Seventy-six percent reported a
diagnosis of MF (n=16), 14% reported a diagnosis of SS, and 10% did not report a diagnosis.
Average disease length was 10 years (SD = 9.50), with a range of less than 1 year to 31 years.
The 14 participants who reported stage of diagnosis indicated that they were stage IA (n =8,
38%), stage IB (n=4, 19%) and stage IIB (n=2, 10%).
Based on qualitative analysis, 43 of the 60 codes developed from the coding scheme reached
agreement of a Cohen’s kappa at or above 0.65. Saturation was reached after 15 patients,
suggesting an adequate sample size. Thematic content analysis identified six major code
groupings (treatment, impact on daily activities, emotional, social, coping and management, and
symptoms/symptom burden), which were used to generate a final of 31 items to create the
preliminary version of the MF/SS-CTCL QoL.

Cognitive Debriefing

Using the same online research platform (ORE) and participant inclusion criteria from the
concept elicitation phase, the revised items were administered to a sample of participants.
Although a partnership between PatientsLikeMe and the Cutaneous Lymphoma Foundation (a
patient-advocacy group) was made to help with patient recruitment, there was substantial overlap
in participants across stages of the research study due to difficulty recruiting patients with these
rare diseases. Participants were asked to complete the preliminary items and to provide specific
quantitative and qualitative feedback regarding clarity/semantic ambiguity and understanding,
relevance, and adequacy of each item and the response options.

Results from Cognitive Debriefing

Forty-two participants took part in cognitive debriefing. Approximately half of the participants
were men (51%) with an average age of 62 (SD = 14.13, range 31-101). The majority reported a
diagnosis of MF (85%). Disease stage ranged from IA to IVA, with most participants reporting
stage IA (39%) or stage IB (17%).

Based on quantitative and qualitative cognitive debriefing results, changes were made to the
instrument; items were removed, items were revised to improve clarity, a response option was
added for patients who were in remission, and the recall period was changed from seven days to
four weeks. The final MF/SS-CTCL QoL contained 14 items.
Psychometric Evaluation

Participants
Patients were eligible to participate if they were members of the online community (PatientsLikeMe), adults (18 years of age or older), and reported a diagnosis of MF or SS. Participants were recruited through the PatientsLikeMe online community with support from the Cutaneous Lymphoma Foundation.

Data Collection
Following consent, eligible participants completed a demographic survey, the preliminary MF/SS-CTCL QoL, and the Skindex-29\(^7\) through the online research platform. Additionally, participants were asked to complete a second administration of the MF/SS-CTCL QoL five days later to evaluate stability of item functioning.

Measures
The Skindex-29\(^7\) is a commonly used and valid 29-item self-report measure that evaluates health-related QoL. Specifically, the Skindex-29 covers facets of QoL such as emotional functioning, physical functioning, and symptoms with a four-week recall period. The preliminary 14-item version of the MF/SS-CTCL QoL required patients to rate their impairment in health-related QoL over the last 4 weeks using a 1 (not at all/never) to 5 (very much/always) Likert-type rating scale. Four of the items also include a sixth response option, “Does not apply (I don’t have symptoms right now).”

As part of the current study, participants also completed a brief demographics survey, asking them to provide information about their sex, age, race, ethnicity, diagnosis, and stage of their diagnosis.

Psychometric Evaluation Procedures
The empirical evaluation included determining: (1) the optimal item response theory (IRT) measurement model; (2) item fit; (3) unidimensionality; (4) rating scale performance; (5) reliability; (6) test information (precision); (7) person-to-item map; (8) convergent and discriminant validity; and (9) presence of bias via differential item function. Analyses were performed in SPSS version 24 and WINSTEPS version 3.74.0.
**Results**

**Participants**

A total of 126 patients completed the survey, and 66 (52%) patients completed the second administration of the survey. Most participants were non-Hispanic (91%), White/Caucasian (86%), and female (59%). Participants ranged in age from 22 to 86, with an average age of 59 years ($SD = 13.5$). Ninety-four percent reported a diagnosis of MF (n=118), and 6 percent (n=8) reported a diagnosis of SS. Participants indicated that they were stage IA (n=56, 44%), stage IB (n=24, 19%), or stage II or above (n=22, 17%), and 19 percent (24) did not know or report a stage. Average disease length was 8 years ($SD = 7.6$), with a range of less than 1 year to 35 years.

**Item Descriptive Statistics**

Item descriptive statistics are presented in Table 1.

<table>
<thead>
<tr>
<th>Item</th>
<th>Min</th>
<th>Max</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. In the past 4 weeks, how much did you worry that your mycosis fungoides or Sézary syndrome may get worse?</td>
<td>1</td>
<td>5</td>
<td>2.74</td>
<td>1.26</td>
</tr>
<tr>
<td>2. In the past 4 weeks, how often did you feel hopeless because of having mycosis fungoides or Sézary syndrome?</td>
<td>1</td>
<td>5</td>
<td>2.01</td>
<td>1.13</td>
</tr>
<tr>
<td>3. In the past 4 weeks, how frustrated were you by the unpredictability of mycosis fungoides or Sézary syndrome?</td>
<td>1</td>
<td>5</td>
<td>2.67</td>
<td>1.36</td>
</tr>
<tr>
<td>4. In the past 4 weeks, how often did you feel depressed or sad because of mycosis fungoides or Sézary syndrome?</td>
<td>1</td>
<td>5</td>
<td>2.15</td>
<td>1.03</td>
</tr>
<tr>
<td>5. In the past 4 weeks, how confident did you feel about managing your mycosis fungoides or Sézary syndrome?</td>
<td>1</td>
<td>5</td>
<td>2.94</td>
<td>1.11</td>
</tr>
<tr>
<td>6. In the past 4 weeks, to what extent were you able to cope with the daily demands (symptom impact and management,</td>
<td>1</td>
<td>5</td>
<td>3.47</td>
<td>1.27</td>
</tr>
<tr>
<td>Question</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. In the past 4 weeks, how severe were your mycosis fungoides or Sézary syndrome symptoms?</td>
<td>1.95</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. In the past 4 weeks, how burdensome was your mycosis fungoides or Sézary syndrome treatment?</td>
<td>2.20</td>
<td>1.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. In the past 4 weeks, how much did your mycosis fungoides or Sézary syndrome limit your daily activities (work inside and outside of the house, self-care such as cooking, cleaning, getting dressed, etc.)?</td>
<td>1.79</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. In the past 4 weeks, how much did mycosis fungoides or Sézary syndrome limit your ability to wear clothes you wanted to?</td>
<td>2.28</td>
<td>1.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. In the past 4 weeks, how often did mycosis fungoides or Sézary syndrome (the condition or associated treatment) leave you too tired to work or do daily activities?</td>
<td>2.11</td>
<td>1.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. In the past 4 weeks, how much did mycosis fungoides or Sézary syndrome negatively affect your relationships with others close to you?</td>
<td>1.73</td>
<td>1.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. In the past 4 weeks, how often did you feel that others do not understand what you are going through with mycosis fungoides or Sézary syndrome?</td>
<td>2.67</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. In the past 4 weeks, to what extent did mycosis fungoides or Sézary syndrome make you feel uncomfortable being around people other than close family and friends?</td>
<td>1.94</td>
<td>1.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Responses of *Does not apply (I don't have symptoms right now)* (0) were marked as missing and removed from analyses when calculating mean and standard deviation.
Identifying the Optimal IRT Measurement Model

Determining the optimal IRT model to calibrate the items was an iterative process based on empirical evidence and substantive rationale [10] [11]. Since items were grouped into two rating scales, frequency and intensity, the Andrich-Grouped Rating Scale Model (G-RSM [11]) and Rating Scale Model (RSM [12]) were considered. Of note, the Partial Credit Model and Generalized-Partial Credit Model were not considered as these models would likely produce unstable estimates due to the number of parameters to be estimated relative to the sample size. To determine whether the rating scales for the intensity items and frequency items could be grouped, respectively, a partial credit model was employed and item characteristic curves (ICCs) were generated. These ICCs were similar within the groups (frequency and intensity). Next, a Global Chi-Squared Test was performed to test whether the G-RSM significantly improved the fit above and beyond the RSM. Results revealed that the G-RSM significantly improved model fit over the RSM, $\chi^2 (3) = 8.78, p = .032$.

Item Fit

Item fit was evaluated by examining item mean square infit and outfit statistics. Two items evidenced infit and outfit statistics above the commonly accepted cut-off of 1.33 [13] and were iteratively removed. Of note, these items still provide useful information about the patient experience and can be used in conjunction with the MF/SS-CTCL QoL global rating (see Multimedia Appendix 1). The remaining items evidenced adequate fit statistics and were retained for further analyses.

Unidimensionality

The assumption of unidimensionality was evaluated via an unrotated principal components analysis on the probability scale residuals in WINSTEPS [14]. Although the eigenvalue
suggested the possible presence of a second dimension (eigenvalue=2.2), evaluation of item content and amount of variance explained by the Rasch measurement model (62.5%) provided support that the 12 items were measuring a unidimensional construct.

**Evaluation of Rating Scale Performance**

Andrich thresholds were examined to further ensure that the rating scales for the set of intensity items and the set of frequency items were performing as expected. Thresholds were ordered, indicating that a higher interference in QoL is required to endorse a high frequency or severity/intensity response category (Figures 1 and 2).

![Category Response Curves for the Frequency Items](image)

**Fig1.** Category Response Curves for the Frequency Items
Legend: This figure depicts the relationship between interference with quality of life and response option selection for the frequency items, whereby the different color curves represent probability of selecting one of the response options. Specifically, “never,” “rarely,” “sometimes,” “often” and “always” are represented by the red, blue, purple, grey, and green curves, respectively. This figure shows that a higher interference in level of quality of life is required to endorse higher frequency.

Fig2. Category Response Curves for the Intensity Items

Legend: This figure depicts the relationship between interference with quality of life and response option selection for the intensity items, whereby the different color curves represent
probability of selecting one of the response options. Specifically, “not at all,” “a little bit,” “somewhat,” “quite a bit” and “very much” are represented by the red, blue, purple, grey, and green curves, respectively. This figure shows that a higher interference in level of quality of life is required to endorse higher intensity.

**Reliability**

Person reliability for the 12-item scale was .87, suggesting that the MF/SS-CTCL QoL is able to discriminate between individuals with low and high levels of interference in their quality of life. Item reliability was .97, which suggests that the sample was large enough to locate items on quality of life. Test-retest reliability ($r = .89, P < .001$), calculated through a Pearson Correlation between MF/SS-CTCL QoL scores at time 1 and time 2 five days later, revealed good to excellent stability.

**Test Information**

A test information curve was generated to evaluate the measurement precision of the MF/SS-CTCL QoL at various levels of the latent trait (quality of life). The test information curve (Figure 3) provides evidence that the amount of interference with QoL was precisely estimated, and that the MF/SS-CTCL QoL is best at differentiating people who have trait levels within about two standard deviations of the mean.
**Fig3.** Test Information Curve

**Legend:** This figure depicts the amount of information (or precision of measurement) that is provided by the MF/SS-CTCL QoL measure across the latent construct of interference with quality of life.

**Person-to-Item Map**

Due to the unique properties of the Rasch model, it is possible to place both persons and items on the same interval level scale or “ruler,” depicted using a person-to-item map. This map can be interpreted as a vertical ruler, with persons (depicted on the left) and items (depicted on the right) ordered in relation to their difficulty or trait level using a scale (logits) with a mean of 0 and
standard deviation of 1. For example, on the MF/SS-CTCL QoL, “In the past 4 weeks, how confident did you feel about managing your mycosis fungoides or Sézary syndrome?” and “In the past 4 weeks, how much did you worry that your mycosis fungoides or Sézary syndrome may get worse?” were found to be easier (i.e., require less impairment in quality of life) to endorse. On the other hand, items near the top of the person-to-item map require a higher impairment in quality of life to endorse (e.g., “In the past 4 weeks, how much did mycosis fungoides or Sézary syndrome negatively affect your relationships with others close to you?” and “In the past 4 weeks, how much did your mycosis fungoides or Sézary syndrome limit your daily activities [work inside and outside of the house, self-care such as cooking, cleaning, getting dressed, etc.]?”). This map is presented below. Overall, examination of the map suggests good person-to-item targeting, or adequate coverage of items across much of the latent trait.
**Fig4.** Person-to-Item Map

**Legend:** This person-to-item map can be interpreted as a vertical ruler, with persons (depicted on the left) and items (depicted on the right). This ruler has a mean of 0, which represents an average amount of interference with quality of life, and standard deviation of 1. Items higher on
the ruler are more difficult or require a greater amount of interference with quality of life to endorse. The map also represents a visual depiction of the MF/SS-CTCL QoL’s coverage across the latent construct of interference with quality of life.

**Convergent and Discriminant Validity**

To evaluate convergent and discriminant validity, a correlation matrix of the MF/SS-CTCL QoL, the Skindex-29, and syndrome stage was constructed. It was hypothesized that the MF/SS-CTCL QoL would be significantly more positively correlated with the Skindex-29, another QoL measure (convergent validity), than syndrome stage (discriminant validity). The correlation between the MF/SS-CTCL QoL and the Skindex-29 \((r = .852, P < .001)\) was significantly greater than the correlation between the MF/SS-CTCL QoL and syndrome stage \((r = .260, P < .001)\), providing support for convergent and discriminant validity.

**Differential Item Function**

Differential item function (DIF) generally occurs when participants with an equal amount of the latent trait (interference in QoL) respond differently to an item [10]. DIF was assessed by gender, age and race (White/Non-White). DIF was considered notable if the DIF contrast estimate was \(>1.0\) logit and significant at alpha=.05. Results revealed that the items did not have DIF at high enough levels to be considered problematic.

**Scoring the MF/SS-CTCL QoL**

A total raw MF/SS-CTCL QoL score is calculated by adding up the patient’s total score from the 12 MF/SS-CTCL QoL items. Table 2 provides scaled scores \((M = 100, SD = 15)\) that correspond to the MF/SS-CTCL QoL total score. Although total raw scores of 10 or 11 are possible due to two items with the response choice “Does not apply (I don’t have symptoms right now)”, these scores should not be interpreted differently from a score of 12. For scoring purposes, “Does not
apply (I don’t have symptoms right now)” is scored as a 0. Further, in order to score the MF/SS-CTCL QoL, each of the 12 items must be completed.

Table 2. Raw to scaled score conversion table.

<table>
<thead>
<tr>
<th>Raw MF/SS-CTCL QoL Score</th>
<th>Scaled Score&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Raw MF/SS-CTCL QoL Score, Continued</th>
<th>Scaled Score&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 or below&lt;sup&gt;b&lt;/sup&gt;</td>
<td>62</td>
<td>37</td>
<td>113</td>
</tr>
<tr>
<td>13</td>
<td>74</td>
<td>38</td>
<td>114</td>
</tr>
<tr>
<td>14</td>
<td>80</td>
<td>39</td>
<td>115</td>
</tr>
<tr>
<td>15</td>
<td>84</td>
<td>40</td>
<td>116</td>
</tr>
<tr>
<td>16</td>
<td>87</td>
<td>41</td>
<td>117</td>
</tr>
<tr>
<td>17</td>
<td>89</td>
<td>42</td>
<td>118</td>
</tr>
<tr>
<td>18</td>
<td>91</td>
<td>43</td>
<td>119</td>
</tr>
<tr>
<td>19</td>
<td>93</td>
<td>44</td>
<td>120</td>
</tr>
<tr>
<td>20</td>
<td>94</td>
<td>45</td>
<td>121</td>
</tr>
<tr>
<td>21</td>
<td>96</td>
<td>46</td>
<td>123</td>
</tr>
<tr>
<td>22</td>
<td>97</td>
<td>47</td>
<td>124</td>
</tr>
<tr>
<td>23</td>
<td>98</td>
<td>48</td>
<td>125</td>
</tr>
<tr>
<td>24</td>
<td>100</td>
<td>49</td>
<td>126</td>
</tr>
<tr>
<td>25</td>
<td>101</td>
<td>50</td>
<td>128</td>
</tr>
<tr>
<td>26</td>
<td>102</td>
<td>51</td>
<td>129</td>
</tr>
<tr>
<td>27</td>
<td>103</td>
<td>52</td>
<td>131</td>
</tr>
<tr>
<td>28</td>
<td>104</td>
<td>53</td>
<td>133</td>
</tr>
<tr>
<td>29</td>
<td>105</td>
<td>54</td>
<td>135</td>
</tr>
<tr>
<td>30</td>
<td>106</td>
<td>55</td>
<td>137</td>
</tr>
<tr>
<td>31</td>
<td>107</td>
<td>56</td>
<td>139</td>
</tr>
<tr>
<td>32</td>
<td>108</td>
<td>57</td>
<td>143</td>
</tr>
<tr>
<td>33</td>
<td>109</td>
<td>58</td>
<td>147</td>
</tr>
<tr>
<td>34</td>
<td>110</td>
<td>59 to 60</td>
<td>154</td>
</tr>
<tr>
<td>35</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>112</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Scaled scores were standardized on the current sample to have a mean of 100 and a standard deviation of 15

<sup>b</sup> While it is possible to obtain a raw score of 10 or 11 due to endorsing “Does not apply (I don't have symptoms right now)” to MF/SS-CTCL QoL items, these scores should be viewed
as equivalent to a 12.

MF/SS-CTCL QoL Interpretation

The qualitative description of scaled scores, provided in Table 3, was based on evaluating the distribution of scaled scores relative to the response categories. Specifically, individuals with scaled scores ranging from 62 to 89 corresponded with an average rating of “1” (Not At All / Never) across the items and were described as having No to Low Interference. Individuals with scaled scores ranging from 91 to 105 corresponded with an average rating of “2” (A Little / Rarely) and were described as having Mild Interference. Individuals with scaled scores ranging from 106 to 117 corresponded with an average rating of “3” (Somewhat / Sometimes) across the items and were described as having Moderate Interference. Individuals with scaled scores ranging from 118 to 133 corresponded with an average rating of “4” (Quite A Bit / Often) across the items and were described as having Substantial Interference. Finally, individuals with scaled scores ranging from 135 to 154 corresponded with an average rating of “5” (Very Much / Always) across the items and were described as having Severe Interference.

Table 3. Qualitative descriptions of MF/SS-CTCL QoL scaled scores.

<table>
<thead>
<tr>
<th>Scaled Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>62 to 89</td>
<td>No to Low Interference</td>
</tr>
<tr>
<td>91 to 105</td>
<td>Mild Interference</td>
</tr>
<tr>
<td>106 to 117</td>
<td>Moderate Interference</td>
</tr>
<tr>
<td>118 to 133</td>
<td>Substantial Interference</td>
</tr>
<tr>
<td>135 to 154</td>
<td>Severe Interference</td>
</tr>
</tbody>
</table>
Discussion

Principal Results

This research utilized a multi-stage instrument development process that incorporated both qualitative and quantitative components, including (1) development of a conceptual model through literature review and input from key opinion leaders, (2) refinement of the conceptual model and generation of preliminary items through concept elicitation with patients, (3) item revisions based on feedback from patients during cognitive debriefing, (4) empirical testing to evaluate psychometric functioning and finalize the MS/SS-CTCL QOL. Results provide strong support for reliability and validity of the MS/SS-CTCL QOL. Specifically, results indicate that the rating scale was functioning as expected, and the 12-item MS/SS-CTCL QOL exhibited adequate person reliability, excellent test-retest reliability, high levels of measurement precision, good person-to-item targeting, and evidence of convergent and discriminant validity. Items did not evidence significant bias based on gender, age, or race.

The current study employed state-of-the-art modern test theory approaches to instrument development, which are considered the “gold standard” in test construction methodology as they rely on stronger measurement assumptions and produce more reliable results than classical approaches [10] [15]. Further, Rasch/IRT modeling allows for new items to be incorporated into the instrument without having to establish validity of the entire bank. This advantage may be particularly important as treatments improve and disease management changes over time.

Limitations

The sample of patients with MF who participated in this validation study was largely of clinical stage I. Incorporating a greater number of patients who represent the more advanced stages in the item generation process may have resulted in different item content. Consequently, gathering
feedback from patients with more advanced stages will likely be a critical part of future instrument refinement. All data collected from patients during the current study relied exclusively on patient-report, and patient diagnosis and stage could not be verified by a licensed medical professional. Additionally, patients were recruited from the Internet, potentially excluding patients who do not have Internet access or patients whose health or functioning may interfere with their ability to use the Internet. Evaluating the psychometric functioning of this instrument across different settings, such as hospitals and university clinics, is an important next step. Despite the partnership from the Cutaneous Lymphoma Foundation to assist with recruitment of patients with this rare disease, obtaining sample sizes adequate for each phase of measure development was challenging, and the same patients participated in several stages of the development process. Additional studies should be performed to replicate these findings. Future research might also evaluate this instrument’s ability to detect change over time as patient stage, treatment, or health status changes.

Conclusions

The MF/SS-CTCL QoL is the first MF/SS-specific instrument to capture the impact of MF/SS-CTCL on the patient’s health-related QoL. Incorporating the patient perspective throughout the development process likely increased the relevancy of MF/SS-CTCL QoL content for this patient population. The MF/SS-CTCL QoL was developed in partnership with the Cutaneous Lymphoma Foundation with the intention of improving care for MF/SS-CTCL patients. Therefore, the MF/SS-CTCL QoL is free for clinicians, patients, and researchers, and can be downloaded free of charge from the Open Research Exchange.
Acknowledgements

JB, MN, MS, and GS designed the study and provided input throughout the study. JB collected the data. YK provided clinical expertise throughout the study and assisted with finalization of the instrument. SM, RB, and JB analyzed portions of the data. SM wrote the manuscript along with contributions by all authors. All authors read and approved the final manuscript. The authors would like to thank Jean-Philip Okhovat, MD, MPH for his contribution to the manuscript.

This research was funded by Actelion US, Inc. Authors MN, MS, and GS were employees of Actelion US, Inc. at the time this research was conducted.

Conflicts of Interest

This research was funded by Actelion US, Inc. Authors MN, MS, and GS were employees of Actelion US, Inc. at the time this research was conducted.

Abbreviations

MF Mycosis Fungoides
Multimedia Appendix 1: MF/SS-CTCL QoL Instrument
References


