Reliability and Validity of the Ocular Surface Disease Index

Rhett M. Schiffman, MD, MS; Murray Dale Christianson, MD, FRCSC; Gordon Jacobsen, MS; Jan D. Hirsch, PhD; Brenda L. Reis, PhD

Objective: To evaluate the validity and reliability of the Ocular Surface Disease Index (OSDI) questionnaire.

Methods: Participants (109 patients with dry eye and 30 normal controls) completed the OSDI, the National Eye Institute Visual Functioning Questionnaire (NEI VFQ-25), the McMonnies Dry Eye Questionnaire, the Short Form-12 (SF-12) Health Status Questionnaire, and an ophthalmic examination including Schirmer tests, tear breakup time, and fluorescein and lissamine green staining.

Results: Factor analysis identified 3 subscales of the OSDI: vision-related function, ocular symptoms, and environmental triggers. Reliability (measured by Cronbach $\alpha$) ranged from good to excellent for the overall instrument and each subscale, and test-retest reliability was also good to excellent. The OSDI was valid, effectively discriminating between normal, mild to moderate, and severe dry eye disease as defined by both physician’s assessment and a composite disease severity score. The OSDI also correlated significantly with the McMonnies questionnaire, the National Eye Institute Visual Functioning Questionnaire, the physical component summary score of the Short Form-12, patient perception of symptoms, and artificial tear usage.

Conclusions: The OSDI is a valid and reliable instrument for measuring the severity of dry eye disease, and it possesses the necessary psychometric properties to be used as an end point in clinical trials.

Arch Ophthalmol. 2000;118:615-621
SUBJECTS AND METHODS

This study was conducted in compliance with the Code of Federal Regulations for institutional review boards, sponsors, and investigator obligations. Written informed consent was obtained from all patients before enrollment in the study.

SUBJECTS

Patients 18 years or older who had been diagnosed as having dry eye symptoms (International Classification of Diseases, Ninth Revision, code 375.15) at the Henry Ford Health System in the preceding 12 months and who had had symptoms for at least 3 months were recruited for the study. Any patients who had undergone temporary or permanent punctal occlusion were eligible 3 months after the procedure. Control subjects had to have no significant ocular disease other than refractive error and no systemic disease likely to be associated with dry eye. Control subjects were sex-matched and age-matched (± 5 years) to the patients with dry eye disease.

Any patient who had any uncontrolled systemic disease or disability that affected his or her activities of daily living (including ocular allergy, infection, or irritation that was not related to dry eye disease) was excluded from participation in the study. Also excluded were patients who had undergone ocular surgery (including cataract surgery) within the previous 6 months or who were known to have an allergy to any component of any of the agents used in the study, eg, lissamine green, fluorescein, or anesthetic.

All study participants had to have a life expectancy of 1 year or more and “walking-around” visual acuity (with their usual correction) of 20/40 or better in each eye. They had to be English-speaking and cognitively intact (score of ≥20 on the Folstein Mini-Mental State Examination).10

STUDY PROTOCOL

Eligible participants completed a series of questionnaires in which their sociodemographic status, health status, visual functioning, and ocular symptoms were assessed. After completion of the questionnaires, patients underwent a detailed ophthalmic examination. Questionnaires were completed before the examination to ensure that the clinical encounter would not influence the patients’ responses. To determine test-retest reliability of the OSDI, all patients were asked to return in 2 weeks to complete the questionnaire a second time.

MEASUREMENTS

Health Status Measures

Participants completed by self-administration the Short Form-12 Health Survey (SF-12), a measure of general health status;11 the National Eye Institute Visual Functioning Questionnaire (NEI VFQ-25), a questionnaire found to be reliable and valid across several common eye diseases;12 the McMonnies Dry Eye Questionnaire;5 and the OSDI. They were also asked to record their perception of the severity of their dry eye symptoms (as measured by a 9-level subjective facial expression scale).13 They also completed a medical comorbidity survey.

Clinical Measures

Clinical measures included “walking-around” binocular visual acuity according to the Early Treatment of Diabetic Retinopathy Study chart; corneal and conjunctival surface staining using fluorescein and lissamine green, graded according to the Van Bijsterveld14 and Oxford scoring schemes, respectively; tear production according to Schirmer test type 1 (with no anesthesia), Schirmer test type 2 (with nasal stimulation) for those with no tear production on Schirmer test type 1, and basic secretor’s test (with anesthesia); and tear quality (using tear breakup time). The presence of blepharitis, meibomian gland dysfunction, ocular mucin, lid abnormalities, and ocular surface disorders was recorded as well (based on the physician’s judgment). The severity of dry eye disease was assessed in the following 3 ways: by physician assessment, by a composite disease severity score (see below), and by the patients’ frequency of artificial tear use.

The composite disease severity score was created to establish a second measure of severity that was substantially less dependent on the physician’s subjective assessment and easily reproducible. It also adheres to recommendations of the National Eye Institute Workshop by combining traditional clinical measures of dry eye (Schirmer test type 1 and lissamine green staining) with a symptom-based measure (patient perception of ocular symptoms) in the evaluation of dry eye. The criteria for measuring disease severity based on this composite score were set up a priori and were not modified at any point during the study. First, the patient’s findings were designated as normal, mild to moderate, or severe for each individual measure. The cutoff points for the normal and severe groups were based on values suggested by a review of the medical and scientific literature.
However, there was no consensus or consistent trend in the literature indicating appropriate thresholds for distinguishing between patients with mild and moderate values. Therefore, all patients with disease severity graded as greater than normal but less than severe were combined into a single mild to moderate group. The severity designations used for the Schirmer test type 1 were as follows: greater than 10 mm, normal; 6 to 10 mm, mild to moderate; and 0 to 5 mm, severe. The severity designations used for lissamine green staining were as follows: 0 to 0.5, normal; 2 to 6, mild to moderate; and 7 to 9, severe. The severity designations used for the subjective facial expression scale were as follows: 1 to 3, normal; 4 to 7, mild to moderate; and 8 to 9, severe. A value of 1 was assigned to a normal grade for each measure, 2 for a mild to moderate grade, and 3 for severe. The values for the 3 measures were then summed to generate a final severity score: less than 3, normal; 4 to 7, mild to moderate; and 8 to 9, severe.

OSDI Scoring Algorithm

The 12 items of the OSDI questionnaire were graded on a scale of 0 to 4, where 0 indicates none of the time; 1, some of the time; 2, half of the time; 3, most of the time; and 4, all of the time. The total OSDI score was then calculated on the basis of the following formula: OSDI = [(sum of scores for all questions answered) × 100]/(total number of questions answered) × 4).

Thus, the OSDI is scored on a scale of 0 to 100, with higher scores representing greater disability. Subscale scores are computed similarly with only the questions from each subscale used to generate its own score.

STATISTICAL ANALYSIS

The distribution of age, race, sex, income, and educational level between patients with dry eye disease and control patients was compared by t test, χ² test, or Fisher exact test, where appropriate. A P value of .05 was considered significant. Mean functional status scores were computed for all disease severity categories and the control group. The proportion of patients whose data were at the ceiling and floor in each category was also computed.

Factor Analysis

Factor analysis with varimax rotation was used to determine whether items within the OSDI tended to cluster together to create subscales (eg, visual functioning and ocular symptoms). Scores were computed for identified subscales, and these subscales underwent formal reliability and validity assessment along with the total instrument.

Reliability

The reliability of the OSDI questionnaire overall, and for the identified subscales within this study, was computed with Cronbach α, a measure of internal consistency. The test-retest reliability of the instrument was evaluated by computing a random-effects intraclass correlation among patients who completed the OSDI a second time.

Validity

Discriminant validity of the OSDI was assessed by testing for significant differences in OSDI scores by disease severity (based on both the physician’s assessment of severity and the composite disease severity score) by an analysis of variance and the Tukey test for multiple comparisons, which maintains an overall P value of .05. Multiple linear regression was also used to test for an association between disease severity and OSDI scores while adjusting for factors considered likely to influence functional status measures, including sociodemographic factors (age, race, sex, education, income, and employment status) and number of medical comorbidities.

Concurrent validity was computed by correlating OSDI scores with the other measures of health status (McMonies questionnaire, NEI VFQ-25, and SF-12). In addition, correlations with the traditional objective measures of dry eye disease (including lissamine green and fluorescein staining, tear breakup time, Schirmer scores, and artificial tear usage) in the more severely affected (worse) eye were computed with the Spearman coefficient, a nonparametric statistic, as some of the measures were ordinal.

Finally, receiver operating characteristic (ROC) curves were generated to describe the sensitivity and specificity of the OSDI for the diagnosis of dry eye at each OSDI score with the use of both the physician’s assessment of severity and the composite disease severity score. A target enrollment of 150 patients was selected to detect a correlation coefficient of 0.23, detectable with a power of 0.80 at an α level of .05. For estimating test-retest reliability, a sample size of 50 patients was deemed sufficient to provide a 1-sided 95% confidence interval of 0.84 (lower bound) for an intraclass correlation of 0.90. However, to increase the power of the estimate of test-retest reliability, all patients were asked to return for a second test.
The Cronbach \( \alpha \) for the overall instrument and each of the subscales except vision-related function (Table 4). However, when the analysis focused on only patients with Schirmer test type 1 scores less than 10 mm, low to moderate statistically significant correlations were detected for all subscales except vision-related function (Table 4).

In contrast, there was a stronger correlation between OSDI scores and patient-reported variables. All OSDI scores showed moderate correlations with the Patient Perception of Ocular Symptoms that were highly statistically significant \((P < .001)\) (Table 5). There was also a moderate and statistically significant correlation between patient-reported artificial tear usage and both OSDI score and the subscales (Table 2) and once again met or exceeded 0.7.

**VALIDITY OF OSDI**

When patients were grouped according to the physician’s assessment of severity, the mean OSDI total score was 36 in the severe group, 21 in the mild to moderate group, and 10 in the control group (Figure, left; Table 3). The proportion of patients with scores at the ceiling and floor in each category can also be seen in Table 3. All between-group differences were statistically significant \((P < .05)\), Tukey test for multiple comparisons) with the exception of the comparison between the normal and mild to moderate groups for the vision-related function subscale, and the mild to moderate and severe groups for the environmental triggers subscale.

When patients were grouped according to composite disease severity scores, the mean OSDI scores were 36 in the severe group, 18 in the mild to moderate group, and 5 in the control group (Figure, right; Table 3). The proportion of patients with scores at the ceiling and floor in each category can also be seen in Table 3. All group-wise comparisons (between the scores for the normal, mild to moderate, and severe groups) for the total OSDI score and each subscale of the OSDI were statistically significant \((P < .05); \) Tukey test for multiple comparisons).

The OSDI total score and each of the subscale scores were also significantly associated with disease severity (as measured by either the physician’s assessment or the composite score) in multivariate analyses adjusting for sociodemographic factors (age, race, sex, education, income, and employment status) and number of medical comorbidities \((P < .005)\).

In comparing disease severity based on the physician’s assessment with that determined by the composite score, it was found that 130 of 139 patients were graded within 1 level of each other. The weighted \( k \), which is a measure of agreement between 2 rating scales,\(^24\) was 0.61, reflecting good agreement. However, patients’ symptoms tended to be designated as slightly less severe according to the physician’s assessment than they were on the basis of the composite disease severity score.

**COMPARISON OF OSDI SCORES WITH INDIVIDUAL OBJECTIVE CLINICAL MEASURES**

The correlation coefficients between the OSDI and tear breakup time, Schirmer test type 1, fluorescein staining, and lissamine green staining were very low when all subjects were analyzed together (Table 4). However, when the analysis focused on only patients with Schirmer test type 1 scores less than 10 mm, low to moderate statistically significant correlations were detected for all subscales except vision-related function (Table 4).

In contrast, there was a stronger correlation between OSDI scores and patient-reported variables. All OSDI scores showed moderate correlations with the Patient Perception of Ocular Symptoms that were highly statistically significant \((P < .001)\) (Table 5).
the total OSDI score and all of the OSDI subscales (P < 0.001) (Table 5). The mean frequency with which patients used artificial tears ranged from 0 to 9 times a day.

There were also moderate correlations between each of the OSDI scores and the McMonnies Dry Eye Questionnaire; all correlations were highly statistically significant (P < 0.001) (Table 5).

Both the total OSDI score and the scores for the 3 subscales were significantly correlated with the NEI VFQ-25 score (P < 0.001) (Table 5). The negative correlations obtained reflect the fact that higher scores on the OSDI represent greater disease impact, while higher scores on the NEI VFQ-25 reflect less disability. Correlations between each of the OSDI scores and the NEI VFQ-25 questionnaire were stronger than those between the OSDI and either the McMonnies Dry Eye Questionnaire or the patient perception of ocular symptoms.

The OSDI was also compared with the SF-12, a general measure of health status (Table 5). There were low to moderate correlations between OSDI scores and the physical component summary score of the SF-12 (r = −0.39; P < 0.001); however, the correlations between OSDI scores and the mental component summary score were poor.

**OPERATING CHARACTERISTICS OF THE OSDI**

The sensitivity and specificity of the OSDI are shown in **Table 6**. The sensitivity value expresses the proportion of patients above the given threshold who have dry eye, while the specificity value expresses the proportion of patients below the given threshold who do not have dry eye. For the computations presented in Table 6, a threshold was selected that maximized the sum of the sensitivity and specificity values. Alternative thresholds could be chosen that would selectively maximize either the sensitivity or the specificity, depending on how the instrument is to be used. As expected, the performance of the OSDI improved as it was used to discriminate more severe disease. The ROC curves can also be used to describe the performance of a test by plotting the sensitivity vs (1 – specificity) for each possible test result. The area under the ROC curve can be thought of as a summary measure for these curves, with 0.5 indicating that the test is no better than chance at predicting dry eye disease and 1.0 indicating a perfect test for dry eye. The areas under the ROC curve for the OSDI demonstrate good to excellent discrimination with the instrument (Table 6).

In the present study, the OSDI demonstrated both high internal consistency (the Cronbach α for the overall instrument and each of the subscales ranged from good to excellent) and good to excellent test-retest reliability in a large sample of patients with dry eye disease and normal controls. The OSDI also demonstrated excellent validity, effectively discriminating between normal, mild to moderate, and severe dry eye disease as defined by both the physician’s assessment of severity and a composite disease severity score. Moreover, the OSDI demonstrated good sensitivity and specificity in distinguishing between normal subjects and patients with dry eye disease.

One of the assumptions of test-retest reliability assessments is that the subject's condition has remained stable between the time of the first test and the time of the retest. However, the symptoms of patients with dry eye disease typically fluctuate, thus violating this assumption of stability. This may explain the modest decrease in the test-retest reliability compared with Cronbach α and may be unavoidable with this population. Despite this, the test-retest correlation still meets...
or exceeds the level recommended for group comparisons of 0.7.

The OSDI scores correlated well with other eye-specific health status measures, such as the McMonnies questionnaire and the NEI VFQ-25. The correlation was not perfect, however, indicating that the OSDI captures unique aspects of dry eye disease not addressed by these other instruments. This is to be expected, considering the difference in the contents and structure of these different questionnaires. The OSDI is specific for dry eye disease and asks patients the frequency of specific symptoms and their impact on vision-related functioning. The McMonnies questionnaire, although also specific for dry eye disease, was designed as a screening test to discriminate patients with dry eye disease from a normal population and primarily uses dichotomous responses (yes or no) to assess the presence of symptoms. The NEIVFQ-25 surveys general ocular health and is not intended to capture the broad range of symptoms unique to a certain ocular disorder.

When the results from all patients were analyzed together, OSDI scores did not correlate well with traditional objective clinical measures of dry eye (such as the Schirmer test type 1). This is consistent with previous studies that also failed to find strong correlations between objective clinical signs of dry eye and patient symp-

### Table 4. Correlation Between Ocular Surface Disease Index (OSDI) Scores and Individual Clinical Measures in the Worse Eye

<table>
<thead>
<tr>
<th></th>
<th>Tear Break-up Time</th>
<th>Schirmer Type 1 Test</th>
<th>Lissamine Green</th>
<th>Fluorescein</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ρ</strong></td>
<td><strong>P</strong></td>
<td><strong>ρ</strong></td>
<td><strong>P</strong></td>
<td><strong>ρ</strong></td>
</tr>
<tr>
<td>All Patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSDI total score</td>
<td>-0.14</td>
<td>.09</td>
<td>-0.21</td>
<td>.01</td>
</tr>
<tr>
<td>Vision-related function</td>
<td>-0.10</td>
<td>.26</td>
<td>-0.14</td>
<td>.10</td>
</tr>
<tr>
<td>Ocular symptoms</td>
<td>-0.15</td>
<td>.09</td>
<td>-0.21</td>
<td>.02</td>
</tr>
<tr>
<td>Environmental triggers</td>
<td>-0.19</td>
<td>.02</td>
<td>-0.18</td>
<td>.04</td>
</tr>
<tr>
<td>Patients With Dry Eye and Schirmer Test Result &lt;10 mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSDI total score</td>
<td>-0.29</td>
<td>.005</td>
<td>-0.38</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vision-related function</td>
<td>-0.20</td>
<td>.05</td>
<td>-0.26</td>
<td>.01</td>
</tr>
<tr>
<td>Ocular symptoms</td>
<td>-0.27</td>
<td>.01</td>
<td>-0.35</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Environmental triggers</td>
<td>-0.34</td>
<td>.001</td>
<td>-0.38</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

### Table 5. Correlation Between OSDI Score and Other Variables

<table>
<thead>
<tr>
<th>Patient-Reported Variables</th>
<th>Other Questionnaires</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Perception of Symptoms (Subjective Facial Expression Scale)</td>
<td>Artificial Tear Usage</td>
</tr>
<tr>
<td></td>
<td><strong>ρ</strong></td>
</tr>
<tr>
<td>OSDI total score</td>
<td>0.669</td>
</tr>
<tr>
<td>Vision-related function</td>
<td>0.483</td>
</tr>
<tr>
<td>Ocular symptoms</td>
<td>0.590</td>
</tr>
<tr>
<td>Environmental triggers</td>
<td>0.832</td>
</tr>
</tbody>
</table>

*OSDI indicates Ocular Surface Disease Index; NEI VFO-25, National Eye Institute Visual Functioning Questionnaire; and SF-12, Short Form-12 Health Survey.

### Table 6. Sensitivity, Specificity, and AUC of the OSDI

<table>
<thead>
<tr>
<th>Based on Physician Rating</th>
<th>Based on Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>TH</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Normal vs all dry eye</td>
<td>Normal vs severe dry eye</td>
</tr>
<tr>
<td>OSDI total score</td>
<td>15.0</td>
</tr>
<tr>
<td>Vision-related function</td>
<td>12.5</td>
</tr>
<tr>
<td>Ocular symptoms</td>
<td>33.3</td>
</tr>
<tr>
<td>Environmental triggers</td>
<td>12.5</td>
</tr>
<tr>
<td>Normal vs severe dry eye</td>
<td>17.0</td>
</tr>
<tr>
<td>OSDI total score</td>
<td>13.0</td>
</tr>
<tr>
<td>Vision-related function</td>
<td>17.0</td>
</tr>
<tr>
<td>Ocular symptoms</td>
<td>17.0</td>
</tr>
</tbody>
</table>

*AUC indicates area under the receiver operating characteristic curve; OSDI, Ocular Surface Disease Index; and TH, threshold at which combined sensitivity and specificity values were at maximum.
toms. This lack of correlation may be because, among a heterogeneous group of patients with dry eye, these measures lack sufficient sensitivity to capture the full range of ocular surface and tear abnormalities that produce typical dry eye symptoms. However, in certain subsets of patients, these clinical measures may correlate more closely with patient perception of disease severity. In support of this, the present study found that the OSDI correlated moderately with clinical signs among patients with dry eye disease who had tear deficiency.

The OSDI is, therefore, a unique instrument able to assess both the frequency of dry eye symptoms and their impact on vision-related functioning. The OSDI has good to excellent reliability, validity, sensitivity, and specificity. It should prove a valuable complement to other clinical and subjective measures of dry eye disease by providing a quantifiable assessment of dry eye symptom frequency and the impact of these symptoms on vision-related functioning. It also has the necessary psychometric properties to be used as an end point in clinical trials of dry eye disease.

Accepted for publication December 10, 1999.

This study was supported by an unrestricted grant from Allergan Inc, Irvine, Calif. None of the non-Allergan authors have any financial interest in Allergan Inc.

Reprints: Rhett M. Schiffman, MD, MS, Henry Ford Hospital, 2799 W Grand Blvd, Detroit, MI 48202 (e-mail: rmschfmn@aol.com).

REFERENCES