An Objective Approach to Dry Eye Disease Severity

Benjamin D. Sullivan,1 Diane Whitmer,2 Kelly K. Nicols,3 Alan Tomlinson,4 Gary N. Foulks,5 Gerd Geerling,6 Jay S. Pepose,7 Valerie Kosheleff,1 Allison Porreco,1 and Michael A. Lemp1

PURPOSE. A prospective, multisite clinical study (10 sites in the European Union and the United States) evaluated the clinical utility of commonly used tests and tear osmolarity for assessing dry eye disease severity.

METHODS. Three hundred fourteen consecutive subjects between the ages of 18 and 82 years were recruited from the general patient population, 299 of which qualified with complete datasets. Osmolarity testing, Schirmer test without anesthesia, tear film breakup time (TBUT), corneal staining, meibomian dysfunction assessment, and conjunctival staining were performed bilaterally. A symptom questionnaire, the Ocular Surface Disease Index (OSDI), was also administered to each patient. Distributions of clinical signs and symptoms against a continuous composite severity index were evaluated.

RESULTS. Osmolarity was found to have the highest correlation coefficient to disease severity ($r^2 = 0.55$), followed by conjunctival staining ($r^2 = 0.47$), corneal staining ($r^2 = 0.43$), OSDI ($r^2 = 0.41$), meibomian score ($r^2 = 0.37$), TBUT ($r^2 = 0.30$), and Schirmer result ($r^2 = 0.17$). A comparison of standard threshold-based classification with the composite severity index revealed significant overlap between the disease severities of prospectively defined normal and dry eye groups. Fully 63% of the subjects were found to be poorly classified by a best marker of disease severity across normal, mild/moderate, and severe categories. Other tests were found to be informative in the more severe forms of disease; thus, clinical judgment remains an important element in the clinical assessment of dry eye disease. The results also indicate that the transition and progression of dry eye is multifactorial and supports the rationale for redefining severity on the basis of a continuum of disease.5,7-10 In addition, disease expression is known to be variable during the early stages, when compensatory mechanisms may transiently alleviate the effects of environmental stress.6

Dry eye is a multifactorial disease commonly associated with aging, hormonal dysfunction, contact lens wear, systemic drug effects, Sjögren’s syndrome, and refractive surgery.1-5 A general consensus on the mechanistic classification of the disease has defined two main subtypes, aqueous and evaporative dry eye, which loosely correspond to disorders of the lacrimal and meibomian glands, respectively.3 The clinical differentiation of these subtypes is problematic, since many cases of dry eye disease are likely to be a mixture of the two classes.6 Moreover, the underlying mechanisms of the disease are beyond the scope of clinical observation, with androgen deficiency, hyperosmolarity, inflammation, lipid composition, and tear film instability playing definitive roles in the initiation and/or progression of the disease.5,7-10 Most of these measures are specific for one subtype of the disease, yet are used extensively in clinical trials as inclusion criteria and as primary endpoints for determining drug efficacy. These methods also rely on subjective judgment, lack specificity, and are prone to operator-dependent analytical errors.5,12,15

Owing to the variety of disease processes associated with the progression of the disease, dry eye is frequently characterized by conflicting signs. For example, a patient with low aqueous tear production (Schirmer, $<5$ mm) may present with a stable tear film (TBUT, $>20$ seconds). The inverse is also common. The need to find a way to reconcile these conflicts into a single estimate of disease severity is not trivial, because signs carry significant analytical variation, the relative importance of each sign is not clearly established, and signs indicative of subtypes of the disease do not necessarily correlate with overall disease severity.5 In addition, the paucity of data available on the longitudinal or diurnal variation of existing clinical signs may play a role in the lack of correlation between these signs and symptoms of dry eye.14

Unlike tests correlated to subsets of dry eye, osmolarity is believed to be a global indicator of the disease, independent of its etiology.3 In 2006, Tomlinson et al.13 concluded that “the measurement of tear film osmolarity arguably offers the best means of capturing, in a single parameter, the balance of input and output of the lacrimal system. It is clear from the comparison of the diagnostic efficiency of various tests for keratoconjunctivitis sicca, used singly or in combination, that osmolarity provides a powerful tool in the diagnosis of KCS and has the potential to be accepted as a gold standard for the disease.”

From the 1TearLab Corp., San Diego, California; 2Starlab, Barcelona, Spain; the 3College of Optometry, The Ohio State University, Columbus, Ohio; the 4Department of Vision Sciences, Glasgow Caledonian University, Glasgow, Scotland, United Kingdom; the 5Kentucky Lions Eye Center, University of Louisville, Louisville, Kentucky; the 6Department of Ophthalmology, University of Würzburg, Würzburg, Germany; and the 7PeposeVision Institute, Chesterfield, Missouri.

Supported by Alcon Laboratories, Inc. and TearLab Corp.

Submitted for publication February 16, 2010; revised May 23, 2010; accepted June 22, 2010.

Disclosure: B.D. Sullivan, TearLab Corp. (E, I); D. Whitmer, None; K.K. Nicols, TearLab Corp. (I, F); A. Tomlinson, TearLab Corp. (I, F); G.N. Foulks, TearLab Corp. (I, F); G. Geerling, TearLab Corp. (F); J.S. Pepose, TearLab Corp. (I, F); V. Kosheleff, TearLab Corp. (E, I); A. Porreco, TearLab Corp. (E, I); M.A. Lemp, TearLab Corp. (C, I)

Corresponding author: Benjamin D. Sullivan, TearLab Corp., 7360 Carroll Road Suite 200, San Diego, CA 92121; bdsullivan@tearlab.com.
Recently, a technology has become available to allow minimally invasive measurement of tear osmolarity on 50 nL samples. In its 510(k) submission, the TearLab osmometer (San Diego, CA) recorded an average within-run coefficient of variation of 1.47%, which corresponded to a precision of ±4.5 at 308.7 mOsm/L. Accuracy data obtained on 120 samples, performed at three independent physician office laboratories over 2 weeks, demonstrated an $r^2 = 0.9515$ against a laboratory reference vapor pressure osmometer (Vapro 5520; Wescor, Logan, UT). The advent of this technology allows tear osmolarity to be measured easily in a point-of-care setting and enables multisite trials, to investigate tear osmolarity.

The principal purpose of this study was to understand the relationship between common signs and symptoms of dry eye and the severity of disease. Because there is no consensus on a gold standard for disease severity, we applied an independent components analysis (ICA) to create an objective, composite index of disease severity. The ICA-weighted composite produced equal correlation risk for each of the measures and resulted in a continuous description of disease severity, rather than the more familiar, but discrete, clinical grades found in the literature.

**METHODS**

**Clinical Study Design**

Data were obtained from a prospective, multisite clinical study (10 sites in the European Union and United States). Subjects between the ages of 18 and 82 years were included in the study with an average age of 46.3 years. Participants were chosen from the general patient population and from the staff of each clinic. Of the first 314 subjects enrolled, 299 ($n = 218$ female, $n = 81$ male) were used in the analysis, with the remainder disqualified for incomplete case report forms.

Subjects were excluded from the study if they exhibited any active infection of the eye, active ocular allergy, evidence of lid deformity or abnormal lid movement disorder, refractive surgery within 1 year of the study visit, pregnancy or lactation, abnormal nasolacrimal drainage, punctal plug placement within 30 days of testing, or evidence of a systemic disease known to affect tear production. Patients were excluded if they initiated or altered the dose of chronic systemic medication known to affect tear production within 30 days of testing; for instance, initiation or modulation of antihistamines, antidepressants, diuretics, corticosteroids, or immunomodulators was an exclusion criterion. The subjects were required to remove contact lenses at least 8 hours before examination and did not use artificial tears within 2 hours of screening. Informed consent for the research was obtained from the patients or subjects, in accordance with HIPAA regulations and the Declaration of Helsinki. The protocol was approved by the institutional review boards at each site.

Tear osmolarity, slit lamp examinations, Schirmer tests, TBUT, corneal staining, meibomian score, conjunctival staining, and OSDI were performed on both eyes. Osmolarity was measured with a laboratory-on-a-chip, to simultaneously collect and analyze the electrical impedance of a 50 nL tear sample from the inferior lateral meniscus (TearLab Osmolarity System). Slit lamp examinations evaluated the cornea at a magnification of 10× to 16× for the presence of active inflammation or structural change; the iris and anterior chamber for inflammation; and the eyelids for crusts, collarettes, or scales. A 5-minute Schirmer test was performed with sterile strips without anesthetic (Tear Flo; Sigma Pharmaceuticals, Monticello, IA). TBUT was measured by instilling 5 µL of a 2% sodium fluorescein solution and calculating the average of three consecutive breakup times, manually determined with a stopwatch. Corneal staining was evaluated under cobalt blue illumination 2.5 to 3 minutes after fluorescein instillation. Staining amplitude followed the National Eye Institute (NEI)/Industry Workshop scale. For symptom assessment, the OSDI symptom questionnaire was used. Meibomian dysfunction was assessed to grade the quality, expressibility, and volume of gland secretion, according to the Brön/Foukls scoring system. Conjunctival staining was performed 2.5 to 3.0 minutes after 10 µL of a 1% sodium lissamine green dye was instilled. Conjunctival staining amplitude also followed the NEI/Industry Workshop scale. Staff at each clinical site was trained in these methods before initiation of the trial. For each sign, the more severe measurement of the two eyes was used in the analysis of disease severity.

**Mapping of Raw Clinical Data to a Continuous Severity Scale**

To convert the various clinical measurements into a common unit system, an expert panel of clinicians provided a quantitative version of the Dry Eye Workshop (DEWS) severity scale, as shown in Table 1. Based on these breakpoints, tear osmolarity, Schirmer tests, TBUT, corneal staining, meibomian score, conjunctival staining, and OSDI were converted to severity scores ranging between 0 (representing the least evidence of disease) and 1 (representing the most evidence of dry eye). For instance, a Schirmer result of 2 mm would be converted to a severity score of 0.75. Clinical measurements were then normalized by ICA to remove correlation bias and were combined into a composite score. The final composite score was then also scaled between 0 (representing the least evidence of disease) and 1 (representing the most evidence of dry eye). As a control, the clinical trial dataset was replaced with uniformly random values between the minimum and maximum of each clinical sign. The control dataset was mapped according to the severity scale detailed in Table 1 and combined into a composite index. This process was repeated 20 times, and the average correlation coefficients were recorded to discover the expected correlation risk imparted by the composite method.

**Comparison of Dry Eye Classification between Threshold-Based and Continuous Methods**

In the prospective study protocol, classification of dry eye subjects was based on a series of thresholds, without weighting of the severity of each sign. Specifically, the prospective criteria required evidence of symptoms, with an OSDI score ≥5. In addition, at least one eye had to exceed thresholds on two of the five subset signs, chosen from TBUT <7, Schirmer <7, corneal staining >0, conjunctival staining >0, and meibomian score >5. To gauge the agreement between the composite disease severity index and the prospective selection criteria used in this study, histograms of severity index values were plotted for these threshold-derived normal and dry eye groups.

<table>
<thead>
<tr>
<th>Table 1. Modified DEWS Severity Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Severity Grade</strong></td>
</tr>
<tr>
<td>Schirmer test, mm</td>
</tr>
<tr>
<td>TBUT, s</td>
</tr>
<tr>
<td>Staining, NEI/Industry scale</td>
</tr>
<tr>
<td>OSDI</td>
</tr>
<tr>
<td>Meibomian score</td>
</tr>
<tr>
<td>Osmolarity, mOsm/L</td>
</tr>
</tbody>
</table>

Severity grades of 0, 0.25, 0.50, 0.75, and 1.00 are equivalent to the discrete 0, 1, 2, 3, and 4 grading system used in the DEWS report. The ICA coefficients used to remove bias were determined by the mean of the mixed matrix (0.0915, 0.1353, 0.1587, 0.1544, 0.1384, 0.1381, and 0.2056), for osmolarity, TBUT, Schirmer, corneal staining, conjunctival staining, meibomian score, and OSDI, respectively. The continuously mapped, weighted clinical measures were combined into a composite score by the sum of the squared measures divided by the square root of the sum of the weighting coefficients.
Comparison of Individual Signs with Disease Severity

Normal, mild, moderate, and severe groups were segmented by quartile along the disease severity axis. On examination, it was found that there was insufficient resolution to separate mild and moderate patients into two groups, and these subjects were combined into a single category: mild/moderate. The resulting groups contained 75 normal subjects, 149 with mild/moderate dry eye, and 75 with severe dry eye. Raw clinical data were plotted against the composite disease severity index for each of the seven clinical indications. For bilateral measurements, the more severe measurement of the two eyes was reflected on the ordinate of the graphs.

**Table 2. Average Values of Clinical Indications Categorized by Disease Severity**

<table>
<thead>
<tr>
<th>Indication</th>
<th>Normal</th>
<th>Mild/Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osmolarity, mOsm/L</td>
<td>302.2 ± 8.3</td>
<td>315.0 ± 11.4</td>
<td>336.4 ± 22.3</td>
</tr>
<tr>
<td>Schirmer test, mm</td>
<td>19.5 ± 10.4</td>
<td>13.9 ± 9.5</td>
<td>8.2 ± 8.4</td>
</tr>
<tr>
<td>TBUT, s</td>
<td>11.8 ± 6.4</td>
<td>6.1 ± 4.9</td>
<td>2.7 ± 1.5</td>
</tr>
<tr>
<td>Corneal staining</td>
<td>0.4 ± 0.9</td>
<td>1.7 ± 1.9</td>
<td>5.1 ± 4.1</td>
</tr>
<tr>
<td>Conjunctival staining</td>
<td>1.1 ± 1.4</td>
<td>2.6 ± 1.9</td>
<td>5.9 ± 3.6</td>
</tr>
<tr>
<td>Meibomian score</td>
<td>2.6 ± 2.7</td>
<td>5.6 ± 4.7</td>
<td>10.4 ± 5.2</td>
</tr>
<tr>
<td>OSDI</td>
<td>5.5 ± 7.4</td>
<td>21.0 ± 19.2</td>
<td>41.2 ± 21.6</td>
</tr>
</tbody>
</table>

**Figure 1.** Individual signs versus severity. *Vertical dashed lines:* the three quartile-derived groups—normal, mild/moderate, and severe—are demarcated by the severities of 0.20 and 0.35. Within the normal to moderate cohort, only osmolarity showed significant correlation to disease severity. OSDI showed good discrimination for the normal group, but poor discrimination across the remainder of the subjects. The other clinical signs perform well for patients with the most severe disease, but poorly for those in the normal through moderate quartiles.
RESULTS

Disease severity index in normal subjects ranged from 0.100 to 0.203; in patients with mild/moderate disease, it fell between 0.204 and 0.346; and in those with severe disease, it ranged from 0.347 to 0.664. Of interest, none of the subjects was entirely negative across all clinical measures. As such, no subject registered a score of 0. Similarly, patients with the most severe disease did not test the upper limits of the index, which would have required a 0 mm Schirmer strip, 0 second TBUT, complete staining of the cornea and conjunctiva, and a score of 100 on the OSDI.

The average results for each measurement, broken down by quartile-derived severity, are shown in Table 2. Normal values for osmolarity determined by this method (302.2 ± 8.3 mOsm/L) matched those given by earlier studies. Specifically, 9 of the 13 studies cited in the meta-analysis by Tomlinson et al. reported values within 2 mOsm/L of this mean, and the combined average of those studies reported a value of 302 ± 9.7 mOsm/L. A more recent study reported the average in 25 normal control subjects to be 302 ± 18.2 mOsm/L with TearLab osmolarity. (Jacobi C, et al. IOVS 2010;51:ARVO E-Abstract 3381). Similar concordance between normal referent values in the literature and those found in this study was observed for other signs: TBUT (11.8 ± 6.4 seconds vs. 12.8 ± 1.3 seconds), OSDI (5.5 ± 7.4 vs. 4.5 ± 6.6), and Schirmer test (19.3 ± 10.4 mm vs. 20.2 ± 11.3). The average osmolarity in all dry eye subjects in this study, including mild/moderate and severe (322.2 ± 18.8 mOsm/L), was similar, although it was lower than that in Tomlinson et al. (326.9 ± 22.1 mOsm/L), indicating that sample in the present study was most likely from a population with less severe disease than the samples in past studies. TBUT values for severe dry eye (2.7 ± 1.5 seconds) were concordant with those reported by Abelson et al. (2.2 seconds, range 0.9–5.2), and the TBUT average for all dry eye subjects (5.0 ± 4.4 seconds) compared favorably with the 5 second clinical threshold reported for microquantity instillations. A recent study by Miller et al. evaluated symptoms segregated by disease severity and reported an average OSDI in all mild/moderate subjects across baseline and follow-up (20.1 ± 8.4) that was comparable to the average in all mild/moderate subjects in this study (21.0 ± 19.2). Averages across baseline and follow-up for subjects with severe disease in the Miller study (43.1 ± 19.8) were similarly close to the severe group in this study (41.2 ± 21.6). These results confirm that the composite index provided an unbiased method for classifying dry eye patients.

Figure 1 displays the relationship between each of the clinical signs and the composite disease severity index. Each sign is plotted with disease severity on the x-axis and the raw clinical data on the y-axis. The correlation coefficient of each sign is as follows: osmolarity ($r^2 = 0.55$), TBUT ($r^2 = 0.30$), Schirmer ($r^2 = 0.17$), corneal staining ($r^2 = 0.43$), conjunctival staining ($r^2 = 0.47$), meibomian score ($r^2 = 0.37$), and OSDI ($r^2 = 0.41$).

The resulting average correlation coefficients for the random control dataset were: osmolarity ($r^2 = 0.10$), TBUT ($r^2 = 0.15$), Schirmer ($r^2 = 0.08$), corneal staining ($r^2 = 0.09$), conjunctival staining ($r^2 = 0.10$), meibomian score ($r^2 = 0.09$), and OSDI ($r^2 = 0.24$). Therefore, high correlation against the index is not guaranteed, simply because the individual measurements are part of the composite. The one exception was the mapping used for OSDI, which showed a small correlation with the index, even when random noise replaced the clinical values. Removal of OSDI from the composite did not alter the interpretation of the results. Without OSDI as part of the index, the average osmolarity across the three severity grades was 300.8 ± 7.8, 315.5 ± 10.4, and 336.7 ± 22.2 mOsm/L, whereas the correlation coefficients for osmolarity ($r^2 = 0.63$), TBUT ($r^2 = 0.28$), Schirmer test ($r^2 = 0.17$), corneal staining ($r^2 = 0.44$), conjunctival staining ($r^2 = 0.49$), and meibomian score ($r^2 = 0.34$) were essentially unchanged.

Figure 2 exhibits histograms of the composite severity index for the normal and dry eye groups defined by the threshold-based prospective selection criteria. Prospective normal subjects exhibited severities from 0.10 to 0.34, and prospective dry eye patients showed severities ranging from 0.16 to 0.66. Fully 63% of the subjects fell within the overlapping area between severities of 0.16 and 0.34. Examples of prospectively defined normal individuals with clear evidence of disease, as well as prospectively defined dry eye subjects with low disease severities, are shown in Table 3. These data suggest that threshold-based selection criteria failed to properly classify most mild/moderate dry eye subjects.

DISCUSSION

Consistent with its status as a global marker for the disease, tear osmolarity was the only sign that exhibited meaningful correlation across normal, mild/moderate, and severe categories in Figure 1. Unlike subset markers of the disease which had excessive scatter (Schirmer, meibomian scoring, OSDI), bimodality (TBUT), or saturation (corneal and conjunctival staining), the relationship between osmolarity and disease severity was found to be generally linear throughout the dynamic range. These findings are consistent with earlier reports of the marker’s efficacy. In rare cases, there were false negatives produced by osmolarity, which may be the result of reflex tearing, but may also be indicative of the inherent metastability of the tear film within dry eye disease. That said, the authors believe that tear film osmolarity, due to its linearity, objectivity, quantitative nature, and operator independence, is the single best test to augment clinical judgment in the assessment of disease severity.

The dominant feature of dry eye disease in this study was the lack of agreement between individual clinical signs in the progression of the disease. The distribution of signs and symptoms observed in the paper reflect a mix of etiologies and indicates that better terminology is needed to reflect the ways

![Figure 2](image-url)
in which the lacrimal/meibomian/corneal functional unit can become progressively compromised. The data in Figure 1 demonstrate that the clinical presentation of dry eye is that of a continuum. Neither the boundary between normal and mild, nor that between moderate and severe disease is readily identifiable along the severity axis. Conflicting signs, such as those noted in Table 3, are more the norm than the exception. It is therefore evident that dry eye disease does not lend itself to clear categorical distinctions, as the underlying causes of the disease, such as androgen deficiency and inflammation, are broadly related to ocular surface damage and a dysregulation of the tear film.

Especially for the purposes of clinical trials, the data support the use of a composite index for inclusion criteria, since any combination of thresholds would misclassify a reasonable percentage of dry eye patients, as shown in Figure 2. Further, because the longitudinal variability of each of the individual tests has not been established, the values measured at a single snapshot in time may not be consistent during follow-up visits. A composite score tends to insulate estimates of severity against measurement noise.

It is among the strongest validations of the objective approach presented herein that the referent values match data from earlier studies. It could be argued that, given a uniformly distributed population of normal and dry eye subjects, the most valid classification criteria would necessarily synchronize normal means across different clinical measurements. The composite methods used in this analysis have achieved that end without specifically attempting to do so. Therefore, a composite of clinical markers appears to be a valid surrogate for disease severity.

**CONCLUSION**

This study indicates that the progression of dry eye is multifactorial. Many prospectively defined normal subjects showed clear evidence of dry eye. Such conflicts support the rationale for redefining the severity of dry eye on the basis of a continuum of clinical signs and symptoms rather than categorical distinctions. In particular, signs specific for a single mechanistic subtype of dry eye demonstrated poor correlation to disease severity in milder patients. By contrast, tear film osmolarity was found to be the single best marker of disease severity across the normal, mild/moderate, and severe categories.

**Acknowledgments**

The authors thank the principal investigators at the participating sites, including Christophe Baudouin, Jose Benitez-DeCastillo, David Geffen, Joseph Tauber, and Thomas K. Mundorf for collaborating in the study and Michelle Senchyna, Michael Brubaker, and David A. Sullivan for their support.

**References**


