An Assessment of Self-Reported Disease Classification in Epidemiological Studies of Dry Eye


Purpose. To evaluate the reliability of self-reported dry eye disease status and patient-related predictors of misclassification in contact lens wearers.

Methods. Patients completed the Contact Lens Dry Eye Questionnaire (CLDEQ) short form on two occasions. Test-retest reliability of the CLDEQ composite score was determined using the 95% limits of agreement (LoA) and an intraclass correlation coefficient (ICC). The κ statistic was used to determine reliability of disease status and patient-related predictors of misclassification.

Results. The sample included 274 patients. The range of CLDEQ composite scores from both visits was −1.85 to 4.50 and the mean difference between administrations was −0.05 ± 0.75 (P = 0.30). The 95% LoA of the CLDEQ composite score were −1.51 to 1.42 and the ICC was 0.61 (95% confidence interval [CI]: 0.53–0.68). Calculations using the lower limit of the 95% CI showed that three administrations of the survey would be required to obtain a more desirable ICC (0.70). The κ statistic for reliability of dry eye disease classification was 0.58 (95% CI: 0.48–0.67). Logistic regression showed a significant interaction between gender (females) and younger age (P = 0.02) in relation to misclassification of self-reported dry eye disease status.

Conclusions. The reliability of self-reported dry eye disease classification in contact lens wearers is moderate. In epidemiologic studies of factors associated with self-reported disease status, investigators may be well advised to consider using multiple administrations of such outcome instruments and controlling for sociodemographic characteristics to maintain internal validity.

In many recent large-scale, population-based studies, investigators have used either a self-report or symptom-based definition of dry eye disease to examine factors associated with the disease.1-7 Reports of studies using the symptom-based definition have suggested that dry eye disease is prevalent in 0.39% to 33.7% of the population.2,5-10

Approximately 50% of contact lens wearers report experiencing dry eye symptoms at least occasionally.5,11-15 In the United States, there are approximately 35 million contact lens wearers, which suggests that as many as 17 million contact lens wearers experience these symptoms.16 Dry eye symptoms during contact lens wear could lead to a reduction in wearing time or discontinuation of contact lens wear for many patients, although these outcomes are not very well understood. The primary reason for contact lens discontinuation is discomfort and dryness symptoms, although the cause of these symptoms remains elusive.17,18

Because dry eye symptoms affect the wearing prognosis of contact lens wearers, it is important to understand factors that are associated with patient-reported symptoms. The effects of tear film, type of contact lens, and medical, sociodemographic, and other factors associated with these symptoms are not well understood. A key element in measuring factors associated with any disease is the outcome that is used to classify disease status. However, both differential (directional) and nondifferential misclassification can have a significant impact on the internal validity of an epidemiologic study.19-21 Specifically, misclassification of a dichotomous outcome reduces estimates of between-group differences in proportions and biases estimates of risk.

Thus, epidemiologic studies assessing factors associated with self-reported or symptomatic disease must ensure reliable assessments to safeguard from misclassification bias. This is especially true in dry eye disease, in which disease classification often occurs only by self-report and symptoms. However, there is a paucity of research on the reliability of self-reports of dry eye symptoms, especially in contact lens wearers. Further, it is not clear how this reliability might be related to classification of disease status. We recently reported on a survey instrument and scoring algorithm designed to classify the dry eye status of contact lens wearers efficiently (the Contact Lens Dry Eye Questionnaire [CLDEQ]).11 The purpose of this report is to determine the reliability of patient responses to the CLDEQ, quantify misclassification associated with these responses, and determine predictors of misclassification.

Materials and Methods

Patient Sample and CLDEQ Scoring

This research was approved by the Biomedical Institutional Review Board, in accordance with the tenets of the Declaration of Helsinki. Contact-lens-wearing patients from a large ophthalmology clinic were asked to complete the CLDEQ short form on two occasions. The first administration was to recruit and enroll patients in a related contact lens study; the second administration was performed at the time of the study examination. The CLDEQ short form is composed of three primary questions, which we previously showed to be predictors of a doctor’s diagnosis of contact-lens-related dry eye.11 The content of the short form includes two symptom questions (dryness and light sensitivity) and a self-perception question (i.e., do you think you have dry eyes while wearing your contact lenses?). For the two symptom questions, the patient is first asked the frequency of the symptom, which is coded as follows: 1: never; 2: infrequently; 3: occasionally; 4: frequently; and 5: constantly. Unless the patient responds ‘Never’, he or she is then asked about the intensity of each symptom at three time points during the day: (1) within the first 2 hours after inserting the lens, (2) in the middle of the day, and (3) at the end of the day. The intensity responses are coded from 1 (not intense at all) to 5 (very intense). The response options to the self-perception question are yes,
no, and unsure. The responses to the two symptom questions are scored by multiplying the frequency by the average intensity and a constant, which is then used in the algorithm based on parameter estimates from a previous logistic regression.11 These two composite scores are summed to create the scoring index. For this algorithm, patients who respond “yes” to the self-perception question and score greater than 0.05 are classified as having dry eye disease. Patients who respond “no” or “unsure” to the self-perception question and score greater than 1.29 are also classified as having dry eye disease.

Statistical Analyses

All statistical analyses described in this section were performed using SAS, ver. 8.02; SAS Institute, Cary, NC.

95% Limits of Agreement. This method was used to examine the test-retest reliability of the CLDEQ short form scoring index.22 The mean difference of the test-retest values relative to zero represents bias (tested with a one-sample t-test), and the width of the 95% limits of agreement represents the reliability of the outcome. When this analysis is used appropriately, the differences between measures should be normally distributed. Correlations between the difference in scores and the average of the scores, age, and days between survey administrations were determined using Pearson’s correlation coefficient. In general, given our large sample size (n = 274), we had 80% power to find a correlation coefficient as small as 0.17 to be statistically significant (α = 0.05). Therefore, we evaluated the magnitude of the reported correlation coefficient using the following guidelines: a correlation of 0.1 to 0.3 was small, 0.3 to 0.5 was moderate, and greater than 0.5 was large.25 A subgroup analysis was also conducted using a two-sample t-test to determine the impact of patient gender on test-retest reliability of the CLDEQ short form scoring index.

Intraclass Correlation Coefficient. The intraclass correlation coefficient (ICC) was also used to examine the test-retest reliability of the CLDEQ scoring index, and the 95% CI associated with the ICC is also reported. It has been recommended that an ICC should exceed 0.90 if a technique is to be used for individual assessments in clinical practice and 0.70 for discriminating among groups in research.24 We also determined the number of administrations (m) required to obtain a desired ICC (P) using the lower limit of the 95% CI according to the method of Shrout and Fleiss.24

Kappa (κ) Statistic. The simple κ statistic and 95% CI were used to examine the test-retest reliability of dry eye disease status based on the scoring algorithm. Kappa values were classified for reference as follows: less than 0.00 showed poor reliability, 0.00 to 0.20, slight reliability; 0.21 to 0.40, fair reliability; 0.41 to 0.60, moderate reliability; 0.61 to 0.80, substantial reliability; and greater than 0.80, near perfect reliability.25,26 A subgroup analysis was also conducted to determine the effect of days between survey administrations, age, and gender on the test-retest reliability of individual CLDEQ items. The days between survey administrations and age subgroups were determined by using the median values for each of these outcomes.

Logistic Regression. Logistic regression was used to examine patient-related predictors of misclassification of disease status. Predictors examined included gender, marital status, income, age, race, educational status, and time between survey administrations. All predictors were considered for inclusion in multivariate modeling; however, in the final multivariate model, variables were significant only if P < 0.05. Odds ratio (OR), 95% CI, and probability are reported. Multivariate models were determined by a backward-selection technique, and all variables in these models were examined for interactions. Significant predictors from the multivariate models are reported.

RESULTS

Two hundred seventy-four patients were included in this study. The average age of the sample was 30.9 ± 10.8 years (67.5% female). All were current contact lens wearers. The mean (±SD) CLDEQ index score from the first visit was 1.02 ± 0.80 (range = −0.74 to +4.50); the mean CLDEQ index score from the second visit was 0.98 ± 0.88 (range = −1.83 to +4.50). The mean difference between administrations was −0.05 ± 0.75, which displayed a normal distribution (Fig. 1) and did not significantly differ from zero (t = −1.02, P = 0.30). The 95% limits of agreement associated with test-retest reliability were −1.51 to 1.42 units (Fig. 2). There was no correlation between the difference in CLDEQ score and the average index score (r = 0.13, P = 0.04), age (r = −0.02, P = 0.79) or days between survey administrations (r = −0.04, P = 0.51). The mean difference between CLDEQ index scores was −0.08 ± 0.55 in the men and −0.03 ± 0.83 in the women. These mean differences did not differ significantly (t = 0.57, P = 0.57). The ICC for test-retest reliability of patient responses to the CLDEQ index score was 0.61 (95% CI: 0.53–0.68). According to the analysis of Shrout and Fleiss,24 to obtain an ICC of 0.70, three administrations of the CLDEQ would be required.

Table 1 presents the overall test-retest reliability of dry eye classification for the sample. A significant percentage (P < 0.0001) of individuals misclassified their dry eye status, with significantly more patients going from a positive dry eye status to a negative dry eye status (15.9%) than the converse (7.3%). The simple κ statistic associated with the test-retest reliability for these data was 0.58 (95% CI: 0.48–0.67). The simple κ statistic for females was 0.50 (95% CI: 0.37–0.62) and for males was 0.69 (95% CI: 0.53–0.84). The group was stratified based on the median age (26 years) and the κ statistic for the younger subgroup was 0.56 (95% CI: 0.42–0.69) and for the older group was 0.59 (95% CI: 0.45–0.73).

Univariate analyses from logistic regression models shown in Table 2 indicate that no patient-related variables were significantly related to dry eye misclassification. Although not significant in the univariate model, there was an indication of gender-related differences in misclassification, with the women showing a slight predominance in this regard (OR = 1.88, 95% CI: 0.95–3.70). In multivariate modeling, there was a significant interaction between gender and age (P = 0.02), although no other sociodemographic factors were related to
misclassification. In women greater than or equal to 45 years of age, the odds of misclassifying dry eye status was 1.15 (95% CI: 0.39 – 3.37). In women less than 45 years of age, the odds of misclassifying dry eye status was 2.50 (95% CI: 1.20 – 5.20), and in older men it was 6.73 (95% CI: 0.86 – 52.77).

DISCUSSION

To our knowledge, there have been no studies that have evaluated the reliability of symptom reporting in contact lens wearers, although studies commonly report the high frequency of dry eye symptoms in this group.12,14,27 Further, there have been no reports that have evaluated the ability of patients to report dry eye disease status reliably. In this study the reliability of patients’ responses on this contact-lens–related dry eye questionnaire was moderate. The overall CLDEQ index reported showed fair-to-moderate reliability, with both the 95% limits of agreement analysis and the ICC and the κ statistic associated with the cutoff point classifying dry eye status was moderate. A comparison of gender-specific κ statistics showed that women were less reliable in reporting dry eye status than men. This was confirmed in logistic regression models, which showed that, although neither gender nor age was necessarily significant alone in predicting misclassification, there was a significant interaction between the two. In this regard, a young woman was more likely to misclassify her dry eye status than older women or men in general. There were no other socio-demographic factors associated with dry eye misclassification.

There are several potential explanations relating to these findings regarding reliability and dry eye disease classification. First, this disease and symptoms relating to it may be variable, and we may indeed be measuring this variability. Thus, contact lens wearers may have discomfort that they describe as “dry eye,” but it may not be consistent. If this notion of measuring disease variability is true, it gives some indication of the responsiveness of the instrument to this variability. The finding that young women tend to change their self-reported disease classification is also an interesting one. Various reports have suggested that a woman's endogenous hormonal status may play a role in dry eye disease, and these results could support the suggestion that monthly hormonal status in women may be associated with variability in dry eye disease in this subgroup. Future studies should address the relation between monthly hormonal status and variability in dry eye disease in women of child-bearing age.

Sometimes multiple administrations of screening or diagnostic tests are used to increase test sensitivity or specificity or to help reduce misclassification of disease status. These problems have been noted previously in epidemiologic studies evaluating self-reported occupational hazard exposures,20–21 oral contraceptive use,32 and cancer risk studies.33–35 The effect of misclassification is usually a bias in the measure of effect (i.e., the odds ratio).

In another recent study, the reliability of individual dry eye symptoms in a sample of non-contact-lens–wearing patients with dry eye was also shown to be moderate.36 In that study, the weighted κ statistic for the frequency of dryness symptom was 0.62, which is similar to our result for self-reported dry eye classification. Dryness is a key symptom in dry eye disease, and these studies have shown the reliability of patient reports of dryness to be moderate. The Ocular Surface Disease Index (OSDI; Allergan, Inc., Irvine, CA) is a questionnaire that was designed to assess symptoms of ocular irritation consistent with dry eye disease and their impact on vision-related functioning.37 It is a 12-item questionnaire made up of three subscales: vision-related functioning, ocular symptoms, and envi-

<table>
<thead>
<tr>
<th>Table 1. Test-Retest Data Associated with the CLDEQ Classification Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dry Eye Status: First Administration</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Yes</strong></td>
</tr>
<tr>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Dry Eye Status: Second Administration</strong></td>
</tr>
<tr>
<td><strong>Yes</strong></td>
</tr>
<tr>
<td><strong>No</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
| Data are the number of subjects with the percentage of the total group in parentheses.
vironmental triggers. The overall reliability of the OSDI was good (ICC = 0.82), as was each of the subscales (ICC = 0.70 – 0.81). It is not clear how patients completing the OSDI would perform in reliably classifying dry eye disease status. Although the National Eye Institute Visual Function Questionnaire (NEI-VFQ) is not dry eye symptom specific, its reliability has been examined in a sample of dry eye patients. The overall NEI-VFQ index showed good reliability (ICC = 0.88, 95% limits of agreement = −9 to +8 units on a 100-unit scale). However, it is not likely that this instrument could serve as an outcome for classification of dry eye disease status, given the generic nature of its content.

In summary, the reliability of patients’ self-report of dry eye disease status through the use of a screening questionnaire is moderate in contact lens wearers. It could be important in epidemiologic studies of tear film and other factors potentially associated with symptomatic dry eye to consider multiple administrations of such a survey to ensure appropriate classification of disease status. Further, it may be important to control for demographic or other variables in relation to the dry eye status outcome to ensure that the sample is internally valid.

References


Table 2. Univariate Factors Associated with Dry Eye Misclassification

<table>
<thead>
<tr>
<th>Variable</th>
<th>Misclassified</th>
<th>Not Misclassified</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>30.1 (9.7)*</td>
<td>51.1 (11.1)*</td>
<td>0.99 (0.96 – 1.02)</td>
<td>0.52</td>
</tr>
<tr>
<td>Time between Survey Administrations (d)</td>
<td>46.3 (34.7)*</td>
<td>38.3 (33.9)*</td>
<td>1.01 (1.00 – 1.02)</td>
<td>0.12</td>
</tr>
<tr>
<td>Gender (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>22.4</td>
<td>35.3</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>77.6</td>
<td>64.7</td>
<td>1.88 (0.95 – 3.70)</td>
<td>0.07</td>
</tr>
<tr>
<td>Marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner</td>
<td>36.2</td>
<td>37.0</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>No Partner</td>
<td>63.8</td>
<td>63.0</td>
<td>1.04 (0.57 – 1.89)</td>
<td>0.91</td>
</tr>
<tr>
<td>Race (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>81.0</td>
<td>78.7</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Non-white</td>
<td>19.0</td>
<td>21.3</td>
<td>0.87 (0.41 – 1.80)</td>
<td>0.70</td>
</tr>
<tr>
<td>Income (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$&lt;$30K</td>
<td>36.2</td>
<td>36.6</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>30–74,999K</td>
<td>39.7</td>
<td>36.2</td>
<td>1.11 (0.57 – 2.17)</td>
<td></td>
</tr>
<tr>
<td>75,000K</td>
<td>24.1</td>
<td>27.2</td>
<td>0.90 (0.42 – 1.91)</td>
<td>0.85</td>
</tr>
<tr>
<td>Education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS Grad</td>
<td>19.0</td>
<td>22.7</td>
<td>0.86 (0.35 – 2.14)</td>
<td></td>
</tr>
<tr>
<td>AS or BS</td>
<td>60.3</td>
<td>56.0</td>
<td>1.11 (0.53 – 2.32)</td>
<td>0.80</td>
</tr>
<tr>
<td>Masters +</td>
<td>20.7</td>
<td>21.3</td>
<td>Reference</td>
<td></td>
</tr>
</tbody>
</table>

* Mean ± SD.