The Reliability and Validity of McMonnies Dry Eye Index

Kelly K. Nichols, OD, MPH, PhD, Jason J. Nichols, OD, MS, MPH, and G. Lynn Mitchell, MAS

Objective: The purpose of this report was to investigate the psychometric properties of the McMonnies questionnaire for dry eye disease.

Methods: The instrument was administered to 75 patients with dry eye disease on two occasions. Additional dry eye clinical tests, patient-reported dry eye interviews, and the NEI-VFQ-25 were completed during these examinations. Reliability (internal consistency and test–retest), validity (concurrent and discriminant), and accuracy were assessed for McMonnies Index scores.

Results: McMonnies Index showed poor internal consistency (Cronbach $\alpha = 0.43$) and moderate test–retest reliability (ICC = 0.86, 95% confidence interval 0.76 to 0.90). The mean difference between visits was 0.49 (1-sample $t$ test, $t = 0.92$, $P = 0.36$), and the 95% limits of agreement for test–retest reliability were $-8.6$ to $+9.6$. The instrument also showed fair concurrent validity, correlating only with the NEI-VFQ-25 pain subscale, but better discriminant validity when comparing mild and severe patients. Finally, the McMonnies Index demonstrated fair accuracy in identifying severe dry eye patients as defined by signs and symptoms (area under the ROC curve = 0.65).

Conclusions: The McMonnies Index demonstrates fair reliability and validity as a patient-reported instrument for use in patient care and clinical studies of patients with dry eye disease.

Key Words: dry eye, questionnaire, survey, reliability, validity (Cornea 2004;23:365–371)

The McMonnies questionnaire is often used in the clinical care of patients. The instrument has 14 questions that focus on clinical “risk factors” for dry eye. These domains were derived from the literature and include age, gender, contact lens history, dry eye symptoms, previous dry eye treatments, secondary symptoms (associated with environmental stimuli), medical conditions associated with dry eye syndrome (arthritis, Sjögren syndrome, thyroid disease), dryness of mucous membranes (mouth, throat, chest, or vagina), and medication use. The survey was intended to help detect the presence of dry eye disease and individuals at risk for developing dry eye disease because of their exposures to the “provocative factors” identified by the instrument.

To the best of our knowledge, there have been two reports “validating” the instrument. In one study, age- and gender-specific comparisons were made by comparing women older than 45 years of age, keratoconjunctivitis sicca (KCS) subjects, and normal controls. Although not stated, this is somewhat analogous to an assessment of the instrument’s discriminant (known groups) validity. Significant differences were found when comparing the two groups on dry eye symptoms (including dryness, grittiness, and burning), the frequency of these symptoms, exposure to atmospheric irritants, the use of diuretics, sleeping pills, and tranquilizers, and having a history of digestive problems, arthritis, dry mucosa, thyroid disease, and sleep-related eye irritation. There were no differences in KCS subjects and controls on two dry eye symptoms (soreness and scratchiness), swimming irritation, alcohol exposure, the use of antihistamines or oral contraceptives, and having duodenal ulcer. Discriminant analyses of the instrument yielded a 98% sensitivity and 97% specificity. It has been suggested that these are biased estimates because they were derived from the same data from which the classification procedure was developed. More recently, the McMonnies questionnaire was found to have a sensitivity of 92% and specificity of 93% using the weighted-scale scoring algorithm in a group of 50 women with Sjögren syndrome and 124 women presenting for the correction of refractive error (all women over 45 years of age).

Other than these studies, we are unaware of any reports validating the instrument in the true psychometric sense—in other words, reports examining the instrument’s reliability (internal consistency and test–retest reliability), validity (eg, criterion or construct validity), and accuracy. Once these properties are established, and they are shown to be consistent across studies, then the instrument should be used in epidemiologic studies or clinical trials. The purpose of this report was to evaluate the reliability, validity, and accuracy of the McMonnies Index.
MATERIALS AND METHODS

Patient Recruitment

Patients previously diagnosed with dry eye syndrome (International Classification of Diseases, 9th Edition code 375.15) were recruited to participate in this study. Patients with dry eye disease were recruited to evaluate the reliability and validity of patient-reported outcomes (McMonnies questionnaire and the NEI-VFQ-25) and dry eye clinical tests. In accordance with the tenets of the Declaration of Helsinki, patients signed an Institutional Review Board informed consent document before evaluation. Patients were scheduled for two visits, completed within 21 days, and both visits were encouraged to be at the same time of day. A single examiner performed all of the tests and interviews during the examination. To prevent examiner bias, the visit 1 examination form was not reviewed by the examiner before the second visit. Treatment options, the patient’s dry eye status, and study results were not discussed with the patient until after the second visit. No new medications or treatment regimens were initiated by any patients between visits.

Test Procedures

The specifics regarding the test protocol and procedures for each of the dry eye diagnostic tests have been previously described. Briefly, the order of the testing procedures at all examinations was as follows:

1. McMonnies Dry Eye Questionnaire
2. Patient history (dry eye symptoms and medical history)
   - Dry eye symptoms included dryness, grittiness, soreness, redness, tiredness
   - Frequency response options included never, occasionally (2–3 times per week), frequently (almost every day), and constantly (every day)
3. The National Eye Institute Visual Function Questionnaire (NEI-VFQ-25)
4. Tear meniscus height evaluation
5. Fluorescein tear breakup time
6. Fluorescein staining
7. The Schirmer I test (without anesthetic)/ the phenol red thread test
8. Rose Bengal staining.

McMonnies Index

The McMonnies questionnaire can be found in Appendix 1, along with the appropriate weighting scores for each item. Scores are tabulated using a weighted point assignment “based on clinical experience” where individual questions are summed using weights to obtain an overall “Index” score. The Index score can range from 0 to 45, where higher scores are considered more indicative of dry eye syndrome. A cut-point of greater than 14.5 is recommended for a dry eye diagnosis based on previous sensitivity and specificity estimates.

Statistical Analyses

Reliability

The internal reliability of the McMonnies Index was assessed by Cronbach α using data from visit 1. This analysis for homogeneity is based on the average correlation among the items and the number of items in the instrument. A low α indicates that the items do not come from the same domain, as all items in the Index should be correlated if they measure the same thing. It is recommended that the α values for scales be greater than 0.70 to ensure internal consistency. As stated, Cronbach’s α is dependent on the number of items in the instrument, where the larger numbers of items in the instrument increases the α coefficient. Therefore, we also calculated the average interitem correlation, which is independent of the number of items. This technique is suggested by Cronbach and involves applying the Spearman-Brown formula to the α calculation.

The test–retest reliability of the McMonnies Index was assessed using two methods: the 95% limits of agreement and the intraclass correlation coefficient. When the 95% limits of agreement are examined, the mean of the differences relative to zero represents the bias between visits, and the width of the 95% limits of agreement represents the test–retest reliability of the scale. Second, intraclass correlation coefficients (ICC) and their 95% confidence intervals (95% CI) were calculated to assess test–retest reliability. These values are calculated from the estimates of within- and between-subject errors associated with analysis of variance [where the ICC = (errorbetween/(errorbetween + errorwithin)]. It is generally recommended that the ICC exceed 0.90 if an instrument is to be used on individual patients in clinical practice and that the ICC exceed 0.70 for discriminating among groups of patients in research. However, others recommend that the lower limit of the ICC 95% confidence interval should be at least 0.75.

Validity

All validity assessments were made using data from the first examination, and a P value of 0.05 was considered significant for all hypothesis testing. Construct validation was examined via exploratory factor analysis using varimax rotation. This method is used to examine patterns among correlations within a scale and to test whether these correlations are associated with the anticipated scale structure (ie, the overall McMonnies Index scale). If the items within the McMonnies Index are highly correlated, we would expect the scale to have a strong internal structure constituting one “factor” or construct as previously reported (ie, a dry eye construct).

Concurrent validity (ie, agreement of the Index with a “true” value) was assessed by examining the relationship between the Index and dry eye clinical tests (ie, the Schirmer, phenol red, tear meniscus height, fluorescein staining, and rose bengal staining) using Spearman correlation coefficients because of the nonnormal distributions of test results. Concurrent
validity was also assessed by examining Spearman correlation coefficients between the McMonnies Index and the National Eye Institute Visual Function Questionnaire—25 overall score and pain subscale. 8,18–20

Theoretically, those with more severe disease should report worse scores on a dry eye questionnaire. 21–24 Discriminant validity, a component of construct validity, was assessed by testing for differences in survey scores, artificial tear use, and disease severity using 2-sample t tests. Disease severity was defined based on both signs and symptoms similar to that used in the validation of the Ocular Surface Disease Index (OSDI). 20 For that definition, clinical tests and associated cut-points designating severe dry eye were as follows: at least three symptoms reported at least frequently, a rose bengal score greater than grade 3 on the van Bijsterveld scale, a Schirmer 1 score test less than or equal to 5 mm per 5 minutes of wetting, and a tear breakup time of less than or equal to 5 seconds. If a test was positive according to the aforementioned criteria, a grade 2 was assigned; otherwise, a grade 1 was assigned. The values for the four individual measures were then summed to generate a final severity score (range 4 to 8) where mild to moderate patients were classified with a score of 4 to 5, and severe patients were classified with a score of 6 to 8. Finally, the relationship between age and McMonnies Index was examined using the Spearman correlation coefficient.

Accuracy

The McMonnies Index was assessed for accuracy in detecting disease severity (ie, using the sign and symptom definition for dry eye). Values reported include sensitivity, specificity, and area under the receiver operator characteristic (ROC) curve. The sensitivity and specificity values reported are those for the recommended instrument cut-point of 14.5 for dry eye. 5 Discrimination of the instrument was assessed using the following guidelines for area under the ROC curves: 0.5 indicates no discrimination, between 0.7 and 0.8 indicates acceptable discrimination, between 0.8 and 0.9 indicates excellent discrimination, and greater than 0.9 indicates outstanding discrimination. 25

RESULTS

Patient Characteristics

Seventy-five patients were recruited for this study with a median age of 46.2 years (range 21.4 to 81.0 years), and 70% of the sample was female. The average age of women in the sample was 48.4 ± 17.2 years and of men was 45.1 ± 19.2 years. Ninety-six percent of the sample fulfilled the aforementioned mild-to-moderate criterion for dry eye, and the remaining 36% were classified as severe accordingly.

McMonnies Index Scores and Dry Eye Test Data

The mean value for the Index score from visit 1 was 18.6 ± 6.3 (range 4 to 33), and that for visit 2 was 18.2 ± 6.4 (range 2 to 35). Of the sample, 85.3% were positive for dry eye using McMonnies referent score of 14.5 at visit 1. All dry eye test results were nonnormally distributed (Shapiro-Wilk test, P < 0.05) except the results for the phenol red thread test. The median score for the tear meniscus height evaluation was 0.30 mm (interquartile range 0.20 mm to 0.4 mm); that for the TBUT test was 4.5 seconds (interquartile range 3.5 seconds to 7.0 seconds); that for fluorescein staining was 0.5 stains (interquartile range 0.00 to 1.5 stains); for the Schirmer test it was 14.0 mm (interquartile range 7.75 mm to 19.25 mm); for the phenol red thread test it was 22.0 mm (interquartile range 18.0 mm to 27.25 mm); and for rose bengal staining it was 0.50 stains (interquartile range 0.00 to 1.13 stains). The mean NEI-VFQ-25 overall score was 87.9 ± 8.4 (median 90.2), and the average NEI-VFQ ocular pain subscale score was 69.5 ± 18.7 (median = 75.0).

Reliability

The Cronbach α for McMonnies Index was 0.43, and the average interitem consistency was 0.051, indicating poor internal consistency. The intervisit mean difference for McMonnies Index was 0.49 ± 4.65 (Shapiro-Wilk test, P = 0.06). This value did not significantly differ from zero (1-sample t test, t = 0.92, P = 0.36). The 95% limits of agreement for McMonnies Index were –8.6 to +9.6 units, and the intraclass correlation coefficient for McMonnies Index was 0.86 (95% CI 0.76 to 0.90). Graphic representation of the plots of the difference (visit 2 – visit 1) versus mean [(visit 1 + visit 2)/2] for McMonnies Index can be seen in Figure 1.

Validity

Exploratory factor analysis of McMonnies Index revealed six potential constructs within the instrument. However, none of the factor loadings seemed consistent with any logical construct. For instance, one factor consisted of the question regarding swimming irritation and the question regarding dryness of the nose, mouth, throat, chest, or vagina. A second factor consisted of age, gender, artificial tear use, and a history of thyroid abnormality. A third factor consisted of a question regarding dryness of the mouth, nose, throat, chest, or vagina. A fourth factor consisted of a question regarding dryness of the nose, mouth, throat, chest, or vagina. A fifth factor consisted of a question regarding dryness of the nose, mouth, throat, chest, or vagina. A sixth factor consisted of a question regarding dryness of the nose, mouth, throat, chest, or vagina. The correlations between McMonnies Index score and dry eye clinical test results can be found in Table 1. There were no significant correlations between McMonnies Index score and any dry eye clinical test. The correlation between McMonnies Index and the NEI-VFQ overall score was not significant (Spearman r = −0.11, P = 0.38); however, the correlation be-
tween McMonnies Index and the NEI-VFQ ocular pain sub-scale was significant and in the appropriate direction (Spearman \( r = -0.28 \), \( P = 0.02 \)).

As shown in Figure 2, McMonnies Index was also significantly correlated with age (Spearman \( r = 0.31 \), \( P = 0.006 \)). The mean (± SD) McMonnies Index for patients reporting artificial tear usage was 19.8 ± 5.7, and the mean (± SD) Index score for patients not using artificial tears was 16.9 ± 7.0; these values differed statistically (2-sample \( t \) test, \( t = 1.97 \), \( P = 0.05 \)). The mean Index score when stratified by the sign and symptom criteria was 20.9 ± 6.2 for those with severe dry eye and 17.3 ± 6.1 for those with mild-moderate dry eye under this classification scheme (two-sample \( t \) test, \( t = 2.44 \), \( P = 0.02 \)).

**Accuracy**

The accuracy of McMonnies Index in predicting patients who fulfill the sign and symptom criteria for severe dry eye yielded a sensitivity of 82% and specificity of 36% at a cutpoint of 14.5. The area under the ROC curve was 0.65 (95% confidence interval 0.52 to 0.78).

**DISCUSSION**

Symptom assessment plays a key role in the diagnosis of dry eye.\(^{26,27}\) Yet to our knowledge, there are relatively few dry eye patient-reported instruments that have been psychometrically tested.\(^4,20\) The McMonnies questionnaire is often used in clinical practice and research, although studies examining its psychometric properties are few.\(^2,3\) Bandeen-Roche and co-workers were at the forefront of more systematically characterizing patient symptoms, applying latent class models to a series of six dry eye symptom constructs.\(^{28}\) In that study, they found that four groups, based on number of symptoms and reporting patterns, could be identified: group 1 reported no symptoms, group 2 reported 1 or 2 symptoms (primarily grittiness), group 3 reported multiple symptoms (except crusting), and group 4 reported all symptoms.\(^{28}\) The Ocular Surface Disease Index (OSDI) is another psychometrically tested instrument for dry eye patients consisting of three subscales including a vision-related function scale, an ocular symptoms scale, and an environmental triggers scale.\(^{20}\) The NEI-VFQ-25 and, more specifically, its pain subscale have been used in the assessment of dry eye disease, showing good test–retest reliability and good discriminant validity in identifying more patients with more severe dry eye disease.\(^8\)

The McMonnies Index showed poor internal reliability in terms of both the \( \alpha \) coefficient and average interitem correlation. This low internal consistency indicates that the items in the McMonnies Index measure different dry eye constructs. This was confirmed by factor analysis, which revealed six different constructs, none of which seemed to group logically into any domain. The OSDI was shown to have a high internal consistency overall (Cronbach \( \alpha = 0.92 \)), three factors with moderate to good internal consistency, and good overall test–retest reliability (ICC = 0.82).\(^{20}\) Poor internal consistency has impli-

**TABLE 1. Correlations Between the McMonnies Index and Dry Eye Clinical Tests**

<table>
<thead>
<tr>
<th>Dry Eye Clinical Test</th>
<th>Spearman Correlation Coefficient (P Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tear meniscus height evaluation</td>
<td>0.07 (0.56)</td>
</tr>
<tr>
<td>Tear breakup time test</td>
<td>-0.19 (0.11)</td>
</tr>
<tr>
<td>Fluorescein staining</td>
<td>0.09 (0.47)</td>
</tr>
<tr>
<td>Phenol red thread test</td>
<td>-0.09 (0.47)</td>
</tr>
<tr>
<td>Schirmer test</td>
<td>-0.13 (0.28)</td>
</tr>
<tr>
<td>Rose bengal staining</td>
<td>0.09 (0.45)</td>
</tr>
</tbody>
</table>
McMonnies Dry Eye Questionnaire

Scores for Grading are Located Next to Questions

Age:  □ Under 25 Years  □ 25–45 Years  □ Over 45 Years

Gender:  □ Male  □ Female

1. Have you ever had drops prescribed or other treatment for dry eyes?
   □ Yes (6 points)
   □ No (0 points)
   □ Uncertain (0 points)

2. Do you experience any of the following (if Yes, refer to question #3):
   □ Soreness
   □ Scratchiness
   □ Dryness
   □ Grittiness
   □ Burning

3. How often do your eyes have these symptoms?
   □ Never (0 points)
   □ Sometimes (1 point)
   □ Often (4 points)
   □ Constantly (8 points)

4. Do you regard your eyes as being especially sensitive to cigarette smoke, smog, air conditioning, or heating?
   □ Yes (0 points)
   □ No (2 points)
   □ Sometimes (4 points)

5. Do your eyes easily become red and irritated when swimming in chlorinated fresh water?
   □ Not applicable (0 points)
   □ Yes (2 points)
   □ No (0 points)
   □ Sometimes (1 point)

6. Are your eyes dry and irritated the day after drinking alcohol?
   □ Not applicable (0 points)
   □ Yes (4 points)
   □ No (0 points)
   □ Sometimes (2 points)

7. Do you take:
   □ Antihistamine eye drops
   □ Diuretics (fluid tablets)
   □ Tranquilizers
   □ Oral contraceptives
   □ Medication for duodenal ulcer
   □ Medication for digestive problems
   □ Medication for high blood pressure
   □ Other ______________

   Antihistamines (oral or drop) and/or diuretics
   and/or sleeping tablets and/or tranquilizers
   and/or contraceptives = 2 points total
   Ulcer and/or digestive and/or blood pressure
   medication = 1 point total

8. Do you suffer from arthritis?
   □ Yes (2 points)
   □ No (0 points)
   □ Uncertain (0 points)

9. Do you experience dryness of the nose, mouth, throat, chest, or vagina?
   □ Never (0 points)
   □ Sometimes (1 point)
   □ Often (2 points)
   □ Constantly (4 points)

10. Do you suffer from thyroid abnormality?
    □ Yes (2 points)
    □ No (0 points)
    □ Uncertain (0 points)

11. Are you known to sleep with your eyes partly open?
    □ Yes (2 points)
    □ No (0 points)
    □ Uncertain (0 points)

12. Do you have eye irritation when you wake up from sleeping?
    □ Yes (2 points)
    □ No (0 points)
    □ Uncertain (1 point)

cations associated with using the Index as an outcome; specifically, it reduces the power associated with statistical significance tests (ie, effect sizes), making differences harder to find. Thus, when using the McMonnies Index as an outcome for comparing two treatment groups or longitudinally over time, researchers would potentially need to use large sample sizes. This may not be problematic if the instrument is used to “diagnose” rather than to make comparisons.

The McMonnies questionnaire was shown to have fair test–retest reliability. Theoretically, we would expect patients with a stable dry eye status to provide similar responses on a dry eye questionnaire each time the instrument is administered. However, the 14 questions comprising the McMonnies questionnaire include age, gender, previous dry eye treatments, and medical history questions. None of these types of questions would be expected to change in a 21-day interval, which may explain the moderate ICC value observed in these analyses. However, we believe that the moderately high ICC can be better explained by high variability observed between subjects. This notion is supported by the 95% limits of agreement analysis, which shows much more variability in the test–retest data (about ±9 units on a 45-unit scale). In another study, we found that two refractive error-specific quality of life questionnaires had much better test–retest reliability (approximately ±10 and ±12 units on a 100-unit scale). For both of those instruments, the ICC values were also moderately high, supporting the notion that the instruments had fairly high test–retest reliability. We believe that these data provide a good example of a potential limitation of the use and interpretation of the intraclass correlation coefficient.

The McMonnies instrument showed poor concurrent validity in terms of its lack of correlation with any dry eye clinical tests. This is very similar to results observed with the Ocular Surface Disease Index (OSDI), which showed low correlations with tear breakup time, the Schirmer test, fluorescein staining, and lissamine green staining. These results are not that surprising because previous population-based studies have suggested a poor relation between dry eye signs and patient-reported outcomes. Schein and colleagues found in an elderly sample of 2240 patients no association between the Schirmer test and symptom frequency (individual symptom items, one or more sometimes, or one or more often or all of the time). They also found no statistical relation between rose bengal staining and symptom frequency (one or more symptoms sometimes and one or more symptoms frequently or always). Another population-based study by Hay and colleagues including 341 individuals found no association between the Schirmer test and a group dry eye symptoms measured in terms of frequency. The relation between signs and symptoms in patient-reported outcomes associated with dry eye disease certainly represents a quandary in patient care and clinical research.

Concurrent validity was also assessed by examining the relation between the McMonnies Index scores and the NEI-VFQ overall score and the pain subscale score. Although McMonnies Index did not correlate with the NEI-VFQ overall score in our study, it did correlate with the ocular pain subscale score. In a previous study, the OSDI overall score significantly correlated with the NEI-VFQ overall score and McMonnies Index. Although the correlation between the McMonnies Index and the NEI-VFQ is not reported in their study, one might assume that they were correlated given their mutual high correlations with the OSDI. Again, the McMonnies Index and the NEI-VFQ were not correlated in this study, suggesting that either the OSDI measures a construct not captured by McMonnies Index or disease characteristics from the two study patient samples differed.

Previous investigators have suggested that age and gender are risk factors for dry eye disease. We did not feel that age or gender-specific discriminant comparisons in Index scores were necessarily fair because the scoring algorithm for the McMonnies Index automatically weights women higher than men, and older individuals higher than younger individuals. As expected, the McMonnies Index was significantly correlated with age. The Index has been shown to be sensitive and specific within the older population, again probably because it weights these individuals higher. The McMonnies Index performed fairly in differentiating mild-to-moderate from severe dry eye patients. When the patient sample was stratified by artificial tear usage, the McMonnies Index scores also differed (those with higher scores were artificial tear users); this was a bit surprising because one might think that the artificial tear usage may reduce patient-reported symptoms. However, patients using artificial tears may be those with more severe disease and, therefore, may have higher McMonnies Index scores regardless of their treatment.

The McMonnies Index was found to have poor to fair accuracy in terms of predicting patients with more severe disease. These lower values are likely because we evaluated the Index in a sample consisting only of dry eye patients with varying degrees of disease severity rather than normal controls. In this regard, the Index is faced with discriminating between “shades of gray” (ie, mild versus severe dry eye disease) rather than “black and white” (ie, dry eye versus nondry eye). However, the sensitivities we found with this Index are still better than dry eye clinical tests in predicting more severe dry eye disease. For example, Farris and co-workers found the Schirmer test to have a sensitivity of just 10%, and rose bengal staining to have a sensitivity of 58%, in detecting disease from nondisease. Goren and Goren found the Schirmer test to have a sensitivity of 43%, tear breakup time to have a sensitivity of 36%, and rose bengal to have a sensitivity of 4%. Lucca and co-workers found the Schirmer test to have a sensitivity of 25%, and Schein and co-workers found the Schirmer test...
test to have a sensitivity of 15% and rose bengal staining to have a sensitivity of 30%.31,35

Although the McMonnies questionnaire has been used in clinical practice and research for several years, psychometric data regarding its reliability and validity have been lacking. This study suggests that although the instrument has poor internal consistency, its validity does not suffer too terribly. Other studies are needed to show consistency in these findings before final conclusions are drawn regarding the instrument.

CONCLUSIONS

The McMonnies questionnaire showed poor internal consistency and fair validity and accuracy. Given these attributes, we would not recommend the instrument’s use alone in discriminating among mild/moderate from severe dry eye patients. Other psychometric studies are needed, including dry eye patients and normal controls, before final conclusions can be drawn regarding the instrument.

ACKNOWLEDGMENTS

We would like to thank our sponsors including the National Eye Institute (NEI K23 EY00393-02 and NEI K23 EY13766-01).

REFERENCES